THE PHYSICAL SOCIETY

REPORTS ON PROGRESS IN PHYSICS

VOLUME XV (1952)

EXECUTIVE EDITOR
A. C. STICKLAND

Price to non-Fellows £2 10*s. net* (*Postage* 1*s.* 6*d.*)
Price to Fellows (*one copy*) 27*s.* 6*d. net* (*Postage* 1*s.* 6*d.*)

Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road
London S.W.7
1952

Printed by Taylor & Francis, Ltd.
Red Lion Court, Fleet Street, London, E.C.4

CONTENTS

	PAGE
Preface	iv
The Hydrogen Bond. By L. KELLNER	1
Mass Spectrometry. By K. I. MAYNE	24
Range–Energy Relations. By A. E. TAYLOR	49
The Analysis of High Energy Neutron–Proton and Proton–Proton Scattering Data. By R. S. CHRISTIAN	68
Turbulent Motion. By G. K. BATCHELOR	101
Ferrites. By A. FAIRWEATHER, F. F. ROBERTS and A. J. E. WELCH	142
The Band Structure of Metals. By G. V. RAYNOR	173
Galvanomagnetic Effects in Conductors. By D. K. C. MACDONALD and K. SARGINSON	249
Travelling-Wave Tubes. By R. KOMPFNER	275
Cumulative Index, Vols. 1–15	328

PREFACE

The *Reports on Progress in Physics* is again under the general supervision of the Papers Committee of the Physical Society for the selection of authors and subject matter for each volume of the Reports, and the executive editor of the publication is Dr. A. C. Stickland, Secretary–Editor of the Physical Society.

The members of the Papers Committee were as follows :

Chairman :—Professor S. Devons.

Honorary Papers Secretary :—Dr. H. H. Hopkins.

Ordinary Members :—Professor L. F. Bates, Mr. C. H. Collie, Dr. W. Ehrenberg, Dr. E. P. George, Professor F. L. Jones, Dr. G. O. Jones, Professor H. Jones, Mr. H. L. Kirke, Dr. H. Kuhn, Professor H. S. W. Massey, Professor R. E. Peierls, Dr. W. S. Stiles, Dr. T. Vickerstaff, Dr. H. Wilman and Mr. C. G. Wynne.

Cumulative author and subject indexes of Volumes I–XV of the *Reports* are published in this volume.

A. C. STICKLAND,
Secretary–Editor.

THE HYDROGEN BOND

By L. KELLNER
Imperial College of Science and Technology, London

CONTENTS

	PAGE
§ 1. Introduction	1
§ 2. Intermolecular association	2
§ 3. Intramolecular association	9
§ 4. Quantitative absorption measurements and the mass action law	11
§ 5. The hydrogen bond. Discussion of the forces responsible for the association band	17

ABSTRACT. A survey is given of the infra-red spectra of substances possessing hydroxyl groups and of the position of the O–H valence vibration in the solid, liquid and gaseous state and in dilute solutions. The spectra of water, alcohols and carboxylic acids are discussed separately; and evidence for molecular association is adduced from the experiments on dilute solutions and liquids. The degree of polymerization of various alcohols as a function of the concentration has been obtained from quantitative absorption measurements under the assumption that the mass action law holds. The investigation of the temperature effect of the O–H valence vibration has led to an evaluation of the heat of association between a hydroxyl group and the oxygen of an adjacent group or molecule which is of the order of the values obtained from vapour pressure and cryoscopic measurements. Finally, the nature of the forces producing this association is discussed. It is shown that the van der Waals forces are too small to be solely responsible for this special bond, the so-called hydrogen bond, and that the hydrogen bond must be attributed to a quantum-mechanical effect. An account is given of the possible origins of such a quantum-mechanical binding force.

§ 1. INTRODUCTION

It is well known that compounds possessing a hydroxyl group show a tendency to the formation of molecular aggregates. This tendency becomes apparent *inter alia* in the behaviour of the vapour pressure, in cryoscopic measurements, in the shift of infra-red bands from the value for the gas to lower frequencies for the liquid and solid. The classical case of such a substance is water but the phenomena are just as striking in alcohols and carboxylic acids; they are furthermore not confined to the OH group but have been observed in molecules containing NH and CH linkages. The maximum effect however is obtained between hydroxyl groups and an oxygen atom either in the same or a different molecule. In the first case we speak of intramolecular interaction, in the second of intermolecular interaction. The hydrogen sharing in the effect is called the acceptor, the oxygen the donor. The study of the infra-red spectrum of these substances provides particularly useful evidence for this interaction since the OH vibrations are in a region of the spectrum ($\simeq 3\,600$ cm^{-1}) easily accessible to the observer, since they are comparatively independent of the structure of the compound under investigation on account of the small mass of the hydrogen, and since they provide a direct measure of the forces involved. The infra-red spectrum shows that the OH frequency, which lies at about 3 700 cm^{-1} for gaseous substances, is shifted to 3 300 in the liquid and at the same time greatly broadened.

The half widths change from approximately 40 cm^{-1} to approximately 600 cm^{-1}. From cryoscopic measurements of hydroxyl compounds in solution follows an increase of molecular weight, i.e. a polymerization.

Various attempts have been made to account for the binding energy needed in the formation of the polymers. There are two main approaches to this problem: (i) a connection is sought between the electrostatic, inductive and polarization forces of the liquid or solution and the shift of the bands, (ii) a special bond between hydrogen and the donor, the so-called hydrogen bond, is postulated. In this hydrogen bond, the proton is assumed to form a bridge between two atoms.

In the following, a survey will be given of the infra-red studies of substances capable of forming a 'hydrogen bond' and the hydrogen bond postulate examined in the light of the experimental evidence.

§2. INTERMOLECULAR ASSOCIATION

In vapours the hydroxyl compounds are in the monomeric state with the exception of the fatty acids, while in the liquid state they associate to polymers through the hydroxyl groups, i.e. the hydrogen of one molecule attracts the oxygen of the next, and a chain formation takes place. The vibration of the O–H linkage is affected; the frequency in the gas is higher than in the liquid, that is, the force responsible for the valence bond is weakened at the expense presumably of the force producing the association of molecules. The OH frequency of the monomer is in the neighbourhood of 3 700 cm^{-1} while a very broad band with a maximum at approximately 3 300 cm^{-1} is observed in the associated state. It should be possible to observe intermediate states by dissolving the hydroxyl compound in a nonpolar solvent with which it cannot interact, and it was indeed discovered by Errera and Mollet (1937) and Fox and Martin (1937) that in a dilute solution of alcohols in CCl_4 the broad associating band is shifted towards larger frequencies with a decrease in intensity indicating a change in the state of polymerization. At the same time, a second large band appears near the monomeric frequency which increases in intensity on further dilution while the association band disappears completely for concentrations below 0·02 mole/litre. At this concentration, the substances are therefore entirely dissociated into single molecules. The association band of the aliphatic alcohols in dilution shows two peaks at 3 360 and 3 480 cm^{-1} of which the second is ascribed to the presence of double molecules by Errera and Mollet. The positions and intensities of the neighbouring CH frequencies ($\simeq$3 000 cm^{-1}) are not affected by dilution. The same phenomena have been observed for the overtones of the OH frequency. Particular emphasis however will here be given to the fundamental vibration since the assignment of bands in the region of higher harmonics is less certain as there may be overlapping by overtones of other bands. Examples of the characteristic OH frequencies are given in table 1 together with the molar absorption coefficients.

It will be seen that in the vapour state the OH frequency is still higher than in a very dilute solution; this frequency shift is probably mainly due to the van der Waals forces between the solvent and the solute.

That the state of association is affected by the shape of the molecules taking part in it has been shown by Fox and Martin (1937) in an investigation of the spectrum of phenol, benzyl alcohol, diphenyl carbinol and triphenyl carbinol in dilute solutions. The maximum of the association band is found to vary

between 3480 and 3370 cm^{-1} for these substances; its intensity increases with the wavelength and decreases with increasing number of phenyl groups. In triphenyl carbinol, the association band has shrunk into a small hump on the low-frequency wing of the monomeric band. This result seems to indicate that the bulk of the phenyl groups prevents the OH groups of neighbouring molecules from the close approach necessary for the interaction forces to take effect. Friedel (1951) has made a systematic investigation of this steric hindrance in the isomers of di- and tri-substituted phenols. He worked at concentrations of 0·04 to 0·06 mole/litre where the intermolecular association predominates over the intramolecular bonding (see § 3). In *meta* and *para* substitution the polymeric band appears at 3320–3330 cm^{-1} with an average intensity of $73{\cdot}2 \pm 2{\cdot}1\%$ absorption in addition to the monomeric band at 3570 cm^{-1} while in *ortho* compounds the maximum of the polymeric band occurs at 3460–3410 cm^{-1} with the monomeric band in the same position as before. In the latter case, the intensity is weaker and varies more from one substance to the other, the average being 51 ± 7. In all, 21 substances have been studied showing that the immediate neighbourhood of a substituent (as in *ortho* compounds) proves an obstruction to the interaction of the OH groups.

Table 1. OH Frequencies of Alcohols

Substance	Conc. (mole/litre in CCl_4)	Monomer Frequency (cm^{-1})	Monomer Molar abs. coeff.	Polymer Frequency (cm^{-1})	Polymer Molar abs. coeff.
Ethyl alcohol [1]	3·33	3640	not observed	3350	105
	0·10	3640	not observed	3370	25·2
	0·05	3640	4	not present	
	0·005	3640	15	not present	
Vapour 80° c	0·01	3640	2·35	not present	
Phenol [2]	0·174	3610	130	3390	33·8
	0·087	3610	154	3400	23·8
	0·0145	3610	208	3420	19·5
Vapour [3]		3704		not present	
Solid 25° c [4]		not present		3220	not observed
42° c				3380	not observed
101° c				3420	not observed

[1] Errera and Mollet 1937, [2] Fox and Martin 1937, [3] Williams, Hofstadter and Herman 1939, [4] Brattain 1938.

Water vapour possesses two OH valence frequencies, 3756 and 3654 cm^{-1} of which the first is infra-red active and the second appears in the Raman spectrum. In a highly dilute solution in CCl_4, these frequencies are both observed in the infra-red spectrum; they are displaced to 3705 and 3614 cm^{-1} and the absorption bands show a wing on the high-frequency side which seems to be identical with the envelope of the rotational fine structure lines of the vapour spectrum (Fox and Martin 1940). These bands, which are caused by the oscillations of the single molecules, are replaced in the spectrum of the liquid by a very broad band with a maximum at 3300 cm^{-1} which shifts to even lower frequencies in ice. A new band appears furthermore in the liquid; it lies at 2092 cm^{-1} at 70° c and shifts to higher values at lower temperatures, i.e. in the opposite direction to the 3300 cm^{-1} band. The molecular absorption coefficient of this band rises from 2·8 at 70° c to 7·7 at −7° c. Ice shows furthermore a number of

low-frequency bands which persist into the liquid state and must be attributed to lattice vibrations. The deformation frequency of the H_2O vapour at 1595 cm^{-1} shifts to higher frequencies with a decrease in temperature. The spectrum is set out in table 2 for H_2O and D_2O. It is of interest to consider Lyon and Kinsey's (1942) measurements of the infra-red absorption of water of crystallization in this context. They find two bands at 3598 and 3690 cm^{-1} in beryl which they ascribe to single water molecules; they lie about 10 cm^{-1} below the values in dilute solution (table 2). Fused quartz shows one band at 3672 cm^{-1} and $NaBr.2H_2O$ possesses three frequencies: 3229, 3380 and 3510 cm^{-1}.

Table 2. Spectra of Light and Heavy Water (frequencies in cm^{-1})

	Solid	Liquid	Soln. in CCl_4	Vapour
H_2O	[1]160IR	[1]160IR		
	[2]210R	[2]200R (40°c)	not observed	—
	not observed	[2]320–1020R (40°c)	not observed	—
	[3]1644IR	[3]1642IR	not observed	1595
	[3]2222IR	[3]2149IR		
	[3]2225R	[2]2170R	not observed	—
	[3]3156R	[3]3434IR	[3]3614IR	3654R 3756IR
	[3]3259IR	[2]3448R (70°c)	[5]3705IR	
D_2O	not observed	[1]160IR		
		[2]190R (50°c)	not observed	—
	not observed	[2]250–730R	not observed	—
	not observed	[2]1208R		
		[4]1220IR	not observed	not observed
	not observed	[2]1600R*		
		[4]1480IR	not observed	—
	not observed	[2]2515R	[5]2640IR	[2]2650R
			2742IR	[5]2666IR
				[5]2784IR

*=no definite maximum. R=Raman band, IR=infra-red band, [1] Cartwright and Errera 1936, [2] Cross, Burnham and Leighton 1937, [3] Fox and Martin 1940, [4] Plyler and Williams 1936, [5] Borst, Buswell and Rodebush 1938.

Borst, Buswell and Rodebush (1938) discovered a fine structure in the 3714 and 3705 bands of H_2O in very dilute solutions similar to that in the vapour; their results were however not confirmed by Fox and Martin (1940) who took great pains to keep the optical path in their apparatus dry. They could indeed produce this fine structure by admitting a small amount of water vapour to their spectrometer. Attention must be drawn to the selectivity of the valence vibrations: in H_2O, the symmetrical vibration is Raman-active and the anti-symmetric vibrations infra-red active, although the selection rules for an isosceles triangular molecule permit the appearance of both oscillations in the two spectra. In a dilute solution, however, both frequencies are observed in the infra-red, and they are both active in the infra-red spectrum of D_2O vapour. No theoretical explanation has yet been given for this phenomenon. The intensity of these bands is, as in the case of the alcohols, much enhanced when the substance passes from the gaseous to the liquid state in dilute solution.

It is known from cryoscopic and vapour pressure measurements that the carboxylic acids are associated into dimers even when in the gaseous state. They

dissociate into monomers at temperatures above 170° c; in formic acid for instance only 7% of the molecules are dissociated at 3° c. At room temperature, the acids possess a very broad and intense band with a maximum at 3 000 cm^{-1} which was formerly assigned to the C–H bands in hydrocarbons or alcohols. Moreover, its intensity decreases with rising temperature contrary to the experience with C–H frequencies, so that Gillette and Daniels (1936) came to the conclusion that this broad band in the spectrum of the carboxylic acids is caused by fusion of the O–H association band and of the C–H bands. This assumption was proved by Hofstadter and his collaborators in an investigation of the deuterated acids; they studied the influence of the deuteration of the carboxyl group as well as that of the CH groups on the spectrum at two different temperatures, 24° c and 140° c, where the concentration of the monomers is approximately 0·97. The O–D frequency was found at 2 300–2 350 cm^{-1} in the associated state (24° c) and shifted to 2 660 cm^{-1} in the single molecule while the spectrum of CD_3COOH showed a broad band with a maximum at 3 100 cm^{-1} at the lower temperature and a band of 3 500–3 600 cm^{-1} at the higher temperature. The ratio of the O–H to the O–D frequencies is approximately $\sqrt{2}$. These results were confirmed by Davies and Sutherland (1938) and Buswell, Rodebush and Roy (1938) for acetic acid, trichloro-acetic acid and benzoic acid in dilute solutions. The spectra of the dilute solutions are similar to those of the alcohols: they consist of the sharp monomeric and the diffuse polymeric band, which latter is still quite intense at a concentration (0·0005 mole/litre, table 3) for which it has completely disappeared from the alcoholic spectrum. This fact together with the large shift in the position of the association band shows that the forces producing association are much stronger in the acids than in the alcohols or water. In alcohols the width of the association band is of the order of 300–350 cm^{-1}, in acids it rises to 500–600 cm^{-1}.

The C=O frequency is also affected by the state of polymerization of the carboxylic acids showing that the carboxyl group (i.e. its oxygen atom) is instrumental in bringing about the association between molecules. The shift of the C=O vibration is, however, only of the order of 30 cm^{-1} towards lower frequencies in the transition from the monomeric to the dimeric state.

It will be seen from table 3 that the observations of different investigators check fairly well with the exception of the structure of the association band in Buswell, Rodebush and Roy's observations. The association band shows two groups of doublets of which the lower frequency group is less intense than the higher frequency group.

In the spectra considered so far, association takes place between molecules of the same kind. If, however, this interaction is characteristic for the presence of hydrogen and oxygen atoms, it should also occur when the donor and acceptor belong to different compounds. Experimental investigation has shown this to be the case: association occurs between unlike molecules. If substances possessing a hydroxyl or amino group are dissolved in oxygenated solvents such as ketones, ethers etc., the O–H (N–H) frequency shifts towards lower frequencies as compared with its position in an oxygen-free solvent of the same concentration. Freyman (1937) has reported that the second overtone disappears completely when alcohols are mixed with oxygenated compounds and amines. The association effect has also been observed between the hydrogen of the CH groups in $CHCl_3$ and the oxygen of dioxane and acetone (Gordy 1938) and

between the hydrogen of water and the nitrogen in pyridine (Bosschieter, Errera and Gaspart 1938). When water is dissolved in dioxane and alcohols, the 2130 cm^{-1} band is present down to concentrations of 7% in volume of H_2O in the solvent (Gordy 1936, Williams, Gatica and Gordy 1937). That the interaction takes place between the oxygen of the solvent and the hydrogen of the solute has been proved by Gordy's (1938) study of the displacement of the C–O and C=O bands in mixtures of dioxane and acetone with light and heavy water. The 1125 and 875 cm^{-1} bands of pure dioxane are shifted to 1115 and 865 respectively by the admixture of water while the 1740 (C=O) band of acetone is displaced to 1710, approximately by the same amount as in associated carboxylic acids (table 3). The interaction between the non-polar dioxane and hydroxyl compounds

Table 3. O–H and C=O Frequencies of Carboxyl Acids (in cm^{-1})

Substance	Monomer		Dimer		
	O–H (O–D)	C=0	O–H (O–D)		C=O
CH_3COOH	3640 (139° c)	1770	3125 (24° c)		1739[1]
solution	3520	1767	3000		1720[2]
vapour	3580	not observed	3050	2720	not observed[3]
			2970	2660	
				2580	
solution	3520	not observed	3060	2700	not observed[3]
			2920	2570	
CCl_3COOH					
solution	3500	not observed	3000		not observed[2]
solution	3520	not observed	3060	2700	
			2920	2570	not observed[3]
CD_3COOH	3640 (139° c)	1760	3106 (139° c)		1730[1]
CH_3COOD	2653 (139° c)	1770	2299 (139° c)		1725[1]
CD_3COOD	2660 (139° c)	1760	2326 (139° c)		1740[1]
C_6H_5COOH					
solution	3520	not observed	≃3000 (ill-defined)		not observed[2]
C_6H_5COOH					
solution	3550	not observed	3060	2700	not observed[3]
			2920	2570	
C_6H_5COOD					
solution	—	not observed	3090	2150	not observed[3]

[1] Herman and Hofstadter 1939, [2] Davies and Sutherland 1938, [3] Buswell, Rodebush and Roy 1938.

is particularly marked; this fact together with Gordy and Stanford's (1940) study of the shift in the O–D frequency of heavy methyl alcohol in a variety of oxygenated solvents (table 5) proves that there is no correlation between the displacement of the band and the dielectric constant of the solvent. The interaction is a characteristic effect of the donor; a mixture of H_2O and acetone shows two bands at approximately 3600 and 3500 cm^{-1} while H_2O in pyridine leads to a vibration at 3450. If, however, water is dissolved in an equimolecular mixture of acetone and pyridine, all three bands appear in the spectrum at 3570, 3520 and 3430 cm^{-1} (Bosschieter, Errera and Gaspart 1938). Examples of the interaction between unlike molecules are tabulated in tables 4 and 5. It will be seen that the interaction between unlike molecules is always smaller

than that between like molecules with the exception of the mixture CH_3OD and amines.

The attraction between nitrogen and the hydrogen or deuterium of the hydroxyl group of an unlike molecule seems particularly strong as the H_2O–pyridine mixture also shows a large shift.

Table 4. OH (OD) Frequencies in Different Solvents (in cm^{-1})

Solvent	Solute			
	H_2O	D_2O[2]	CH_3OH[2]	Phenol[3]
Acetone	3620 3490[1]	2610	not observed	not observed
Dioxane	3570[4]	2630	3570	3380
H_2O	3330[4]			
D_2O		2480		
Inert solvent	3705 3614	2784 2666	3690	3580
Ethyl acetate				3400
Pyridine	3450[1]			

[1] Bosschieter, Errera and Gaspart 1938, [2] Gordy 1938, [3] Gordy and Nielsen 1937, [4] Gordy 1936.

Table 5. OD Frequency (in cm^{-1}) of CH_3OD in Different Solvents (Gordy and Stanford 1940)

C_6H_6	2681	Ethers	2558
Ketones	2600	CH_3OD	2494
Aromatic aldehydes	2599	Amines	2479
Aliphatic aldehydes	2570		

Table 6. Fundamental of O–H Frequency ν_0 in Gas Phase and Percentage Frequency Shift $\Delta\nu/\nu_0$ for Hydroxyl Compounds

Substance	ν_0 (cm^{-1})	$\Delta\nu/\nu_0$ for (1)	(2)	(3)	(4)
Aliphatic alcohols	3700	0·02	0·04 0·07	0·02	0·07
H_2O	3756 3654	0·01 0·01	—	0·05	0·08
Phenol	3704	0·03	0·08	0·09	0·09
D_2O	2784 2666	0·01 0·01	—	0·06	0·11
CH_3COOD	2653	not observed	not observed	not observed	0·13
CD_3COOD	2660	not observed	not observed	not observed	0·13
CH_3COOH	3640	0·03	0·18	not observed	0·14*
CD_3COOH	3640	not observed	not observed	not observed	0·15

(1) monomer in inert solvent, (2) polymer in inert solvent, (3) polymer in oxygenated solvent, (4) polymer in liquid.

* O–H band fused with C–H band.

In table 6, a survey is given of the frequency shift from the gaseous to the liquid phase for various hydroxyl compounds. The OH vibration in the gas phase is denoted by ν_0, and $\Delta\nu$ measures the displacement from ν_0. The substances are arranged in increasing order of $\Delta\nu_{liquid}/\nu_0$ and show a progressive weakening of the O–H valence force from alcohols to carboxylic acids. Although the O–D frequencies in D_2O are more diminished than the O–H frequencies of H_2O, the effect goes the opposite way in the deuterated acids. Solution in

oxygenated solvents has in general a greater effect than in inert solvents with the exception of the aliphatic alcohols, but comparison here is not conclusive since the measurements were made with instruments of widely differing resolving powers.

Lüttke and Mecke (1949) carried out a systematic investigation of the influence of various solvents on the O–H vibration of phenol and on its dipole moment. Since the dipole moment of phenol is predominantly due to the dipole moment of the O–H linkage as the very small C–H band moments of the phenyl ring almost cancel each other, it is possible to correlate the changes in the dipole moment of phenol with changes in the O–H band. The dipole moment in eleven different solvents was obtained by measurements of the dielectric constant in dilute solutions of small concentration down to 0·003 mole/litre and extrapolation to infinite dilution. The infra-red absorption of the second overtone of the O–H oscillation was investigated for a concentration of 0·05 mole/litre at two temperatures, 20° c and 50° c. At this low concentration, phenol is almost completely dissociated so that any frequency shift and variation of intensity and width of the O–H band can be attributed to the effect of the solvent on the O–H band. It was found that a shift to lower frequencies was always accompanied by an increased width and decreased total absorption of the band, where the total absorption is obtained by integrating the molar absorption coefficient over the band width. These three quantities move in the reverse direction if the temperature is raised. The dipole moment increases with an increasing effect of the solvent. The results are given in table 6 *a*. It can be seen that four different groups of solvents can be distinguished: in the first group (C_6H_{12} to C_2HCl_5) there is very little interaction between solute and solvent; no intermolecular association takes place.

Cyclohexene, benzene and toluene produce a doubling of the band which is attributed by Lüttke and Mecke to a steric effect; the molecules of the solvent can orientate themselves in two different directions with respect to the O–H dipole. In the benzene derivatives, the molecular plane can be parallel or normal to the direction of the O–H linkage while cyclohexene can turn either its aliphatic side or its double bond towards the hydroxyl group. All these compounds possess π-electrons in the plane of the ring; the interaction between the O–H band and the solvents is ascribed to the relatively large polarizability of the double bonds which is greater for the isolated double bond of cyclohexene and consequently causes a greater effect on the infra-red spectrum and the dipole moment of phenol. C_6H_5Cl on the other hand has a smaller effect although the C–Cl linkage possesses a large dipole moment; dipole–dipole interaction cannot therefore be held solely responsible for the association between solute and solvent.

The third group of solvents (CH_3NO_2, $C_6H_5OCH_3$, $C_6H_5NO_2$) causes a frequency shift of approximately 180 cm^{-1} and a sudden decrease of total absorption; the temperature effect is larger than in the preceding groups. The increased interaction is due to the presence in the solvents of oxygen atoms which act as donors.

Finally, the last group (phenol to acetone) shows very broad and weak bands with an ill-defined maximum and a frequency shift of 200–700 cm^{-1}. All these solvents possess donor groups, and it is of interest to note that the interaction between phenol and aldehydes and ethers respectively is stronger than between phenol and phenol.

§3. INTRAMOLECULAR ASSOCIATION

In molecules possessing a donor and acceptor in adjacent positions, interaction should occur within the molecule with a resultant shift in the frequency of the O–H band. This effect can be best observed in the *ortho* substituents of phenol or benzoic acid which show an O–H band in a position between the monomeric and intermolecular association frequencies. The *meta* and *para* compounds in which the donor and acceptor groups are too far apart to interact do not possess such a band. This intramolecular association is observable with concentrations in which the unsubstituted phenol molecules are completely dissociated; it is

Table 6*a*. Influence of Solvents on the O–H Band at 10 340 cm^{-1} and on the Dipole Moment of Phenol

Solvent	Diel. const. of solvent	Frequency (cm^{-1})		Integral absorption		Half width (cm^{-1})		Dipole moment (e.s.u.)
		20°c	50°c	20°c	50°c	20°c	50°c	20°c
C_6H_{12}	2·02	10 340	10 355	8·0	8·8	100	100	$1·32\times10^{-18}$
C_6H_{10} I	2·22	10 340	10 350	5·6	5·4	230	135	1·79
II		9 950	8 950					
CCl_4	2·24	10 330	10 340	8·2	8·8	110	110	1·40
C_2Cl_4	2·29	10 325	10 330	7·7	8·5	120	120	1·37
C_2HCl_3	3·43	10 305	10 315	7·5	8·5	130	130	1·41
$CHCl_3$	4·82	10 295	10 310	7·7	8·0	155	135	1·44
CS_2	2·64	10 285	—	8·2	—	160	—	1·38
C_2HCl_5	3·82	10 285	10 300	7·7	8·0	160	130	—
C_6H_5Cl	5·69	10 250	10 270	9·3	9·9	210	200	1·45
C_6H_4 I	2·28	10 175	10 200	7·8	8·2	270	290	1·45
II		10 330	10 330					
$C_6H_5CH_3$ I	2·34	10 140	10 170	9·0	9·5	340	355	—
II		10 330	10 330					
CH_3NO_2	39·4	10 160	10 190	4·5	5·6	370	340	—
$C_6H_5OCH_3$	4·37	10 130	10 160	6·2	7·5	430	420	—
$C_6H_5NO_2$	33·7	10 120	10 170	5·5	6·5	370	350	—
Phenol (pure)	13·4	10 120	—	—	—	—	—	2·22
C_6H_5CHO	18·0	9 920	10 000	2·8	4·0	540	500	—
$C_6H_5COCH_3$	18·0	9 870	9 900	2·6	3·1	650	720	—
$(C_2H_5)_2O$	4·33	9 700	—	2·7	—	700	—	2·29
$(CH_3)_2CO$	21·5	9 600	—	3·1	—	750	—	—

easily distinguishable from intermolecular association by the fact that the band intensity does not change with concentration as long as the product of concentration and cell length is kept constant. The strength of the absorption depends therefore entirely on the number of absorbing centres and not on the number of surrounding solvent molecules. On the whole, the intramolecular association bands are much sharper than those found in intermolecular association.

O-chlorophenol and catechol show two bands in the region of the O–H fundamental (Davies 1938) whose first overtones have also been observed (Wulf and Liddell 1935) (table 8). Pauling (1936) suggested that the two bands are due to the co-existence of two stable isomeric forms of the *o*-chlorophenol molecule, a *trans* configuration in which the hydrogen is turned away from the donor (Cl) and which causes the higher frequency band, and a *cis* or associated

configuration in which the O–H linkage is bent towards the donor. The interaction weakens the O–H band and therefore lowers the frequency by about 50 cm^{-1}. From the integrated absorption of these two bands, the relative number of molecules in the two isomeric states can be determined. It is found that at room temperature the ratio of *cis* to *trans* isomers is 10:1 (Pauling 1936) and is lowered to 7·98:1 (Davies 1938) at 73° C: a rise in temperature increases the number of *trans* molecules since they possess a higher potential energy. In catechol, which contains two O–H linkages in adjacent positions, the two bands are of nearly equal intensity as has been observed for the fundamental as well as the first overtone; one of the two O–H bonds is assumed to be in the *cis* position and the other in the *trans* position. The *meta* (resorcinol) and *para* (hydroquinone) compounds show only one band as is to be expected (Wulf and Liddell 1935).

No trace of the monomeric band of 3 600 cm^{-1} can be found in *o*-methylsalicylate ($C_6H_4OH.COOCH_3$) at a concentration of 0·017 mole/litre; on the other hand, the intramolecular association becomes strong enough to shift the O–H band to 3 200 cm^{-1}. This substance exists entirely in the *cis* configuration with the O–H linkage turned towards the double bonded oxygen in the carboxyl group (Fox and Martin 1940). Fox and Martin (1939) also investigated the spectrum of *o*-, *m*-, and *p*-methoxybenzoic acid for various concentrations ($C_6H_4.OCH_3.COOH$). Here, we should expect the *meta* and *para* compounds to show the monomeric and intermolecular association bands of the O–H linkage in the carboxyl group. In addition to these two bands, a third band in an intermediate position should be found in the spectrum of the *ortho* compound proving that intramolecular association between the methoxy group and the hydrogen of the carboxylic group takes place.

Indeed, the spectra of the *meta* and *para* methoxybenzoic acid are found to be identical; they show the monomeric band of carboxylic acids in solution at 3 540 cm^{-1} together with a very broad (600 cm^{-1}) association band at 2 490 cm^{-1}; on this are superimposed the sharp bands of the CH and CH_3 groups due to very strong intermolecular association which persists down to concentrations of 0·002 in the characteristic manner of fatty acids, while the *ortho* compound possesses, in addition to the same two bands, an intramolecular band at 3 360 cm^{-1} (table 7). The intensity of the latter increases at the expense of that of the polymeric band as the concentration decreases while the width of the intermolecular band shrinks simultaneously. We meet here with a co-existence of intermolecular and intramolecular association in the same substance, with a preponderance of the former over the latter. The intramolecular association may be compared with the association between unlike molecules mentioned above where donor and acceptor groups belong to the same molecule. Errera and Mollet (1935) observed that salicylic aldehyde is transparent at 6 620 cm^{-1}, the region of the first overtone of the OH fundamental while its two isomers, *ortho*- and *para*-oxybenzoic aldehyde ($C_6H_5OH.CHO$), show selective absorption at this frequency. The authors took their results to prove chelation between the OH and C=O groups, i.e. the formation of an inner ring with complete quantum-mechanical resonance. But the example of the fatty acids has shown that the OH-association band may be shifted to the position of the C–H frequencies, and indeed a very broad, weak band has been observed in the region of the second overtone at 11 000–9 000 cm^{-1} (Lüttke and Mecke 1950). On dilution of salicylic aldehyde a weak monomeric band occurs indicating that a small proportion of the substance exists in the *trans* form.

In the course of a study of the second overtone at 10 140 cm^{-1} of the hydroxyl groups in seven *ortho*-compounds of phenol, Lüttke and Mecke (1950) have formed the following conclusions: the compounds fall into two classes determined by the nature of the intramolecular donor groups. The substances of the first class (substituent = halogen, OH, OCH_3, CN, C_6H_5) show two sharp bands of a varying intensity ratio corresponding to the presence of both *cis* and *trans* forms of the molecule. In the substances of the second class (substituent = NO_2, CHO, COOR, COR) only one broad, diffuse band is found at a much lower frequency so that it must be concluded that these compounds occur only in the *cis* form, i.e. the intramolecular association is very strong. The authors give a quantum-mechanical explanation of the two effects which will be discussed in a later section.

Gordy's investigation (1940) of the C=O frequency in unsubstituted and *ortho*-substituted compounds has shown that in the case of salicylic acid as well as various other benzene derivatives, the C=O band shifts by about 3–4% towards lower values on substitution, a displacement which corresponds to that observed in carboxylic acids (table 3) from the monomeric to the polymeric state. This work proves that in intramolecular just as in intermolecular association the O–H and C=O bonds are simultaneously weakened.

Table 7. Fundamental O–H Frequencies (in cm^{-1}) in *ortho*, *meta* and *para* Compounds

(1)	(2)	(3)	(4)	(5)
O-chlorophenol	3600	3540	not present	0·01
Catechol	3610	3570	not present	0·01
Methyl-salicylate	not present	3200	not present	0·1
M-, *p*-methoxybenzoic acid	3540	not present	2940	0·002
O-methoxybenzoic acid	3530	3360	2940	0·002

(1) Substance, (2) *trans* frequency, (3) Intramolecular association *cis* frequency, (4) Intermolecular association, (5) Molar concentration (mole/litre).

A comparison of table 7 with tables 1 and 3 shows that the monomeric O–H band appears in the same position for both types of association (aromatic alcohols 3 600–3 610, carboxylic acids 3 520–3 540); the polymeric frequency, however, is higher for intramolecular than for intermolecular association. This result seems to indicate that the acceptor and donor of two different molecules can approach more closely to each other than can those in the same molecule where presumabaly steric hindrance is caused by the benzene ring.

§ 4. QUANTITATIVE ABSORPTION MEASUREMENTS AND THE MASS ACTION LAW

In the first two sections of this report the effect of acceptor–donor association on the O–H frequency in various types of compounds has been discussed. In dilute solution, monomers and polymers, formed by association, are seen to co-exist and the intensities of the respective absorption bands are a measure of their relative concentrations. A knowledge of the variation of the absorption coefficients with concentration and temperature therefore provides a method for the determination of the change in equilibrium between single and multiple molecules and of the heat of association.

Badger and Bauer (1937) were the first to measure the absorption of the sharp 10 256 cm^{-1} band (second overtone of the monomeric O–H band) of light acetic acid as a function of temperature and the amount of gas in the cell. They showed that in the completely dissociated gas the decadic logarithm of the transmission at the band centre is proportional to the number of monomers present. For temperatures below the point of dissociation, the deviation of the absorption coefficient from proportionality with the concentration (or pressure) is due to the formation of dimers. The number and therefore the total pressure of the monomers p_m and of the dimers p_d is obtained from a knowledge of the absorption and the total pressure of the gas in the cell. Since the single and double molecules represent a perfect gas at the pressures used, the mass action law holds and the equilibrium constant K_p could be determined: $K_p = p_m^2/p_d$. If $\log K_p$ is plotted against $1/T$ (T = absolute temperature), the points are seen to lie round the straight line $\log K_p = 11{\cdot}789 - 3\,590/T$ obtained by McDougall (1936) from vapour pressure measurements; the absorption measurements have a lower accuracy than McDougall's values which led to a heat of dissociation of dimer into monomer of 16·4 kcal/mole, i.e. 8·2 kcal/mole for each O–H–O bridge. Similar investigations were carried out by Herman and Hofstadter (1938) with heavy acetic acid on the fundamental O–D dimer band at 2 299 cm^{-1} between room temperature and 145° C. As they were unable to cool the gas down to the temperature of complete association, they used McDougall's value for the degree of dissociation at 38·5° C as reference point and again determined the number of dimers and monomers, i.e. their partial pressures, from the observed absorption at the band centre and the total pressure. To obtain accurate results, the integrated absorption over the whole band should have been measured since the association band is broad; only for sharp bands can the maximum absorption coefficient replace the integration over the frequency range of the band. Herman and Hofstadter found a value of $15{\cdot}9 \pm 0{\cdot}7$ kcal/mole for the heat of dissociation of the dimer, i.e. 7·95 kcal/mole per O–D–O bridge. Their experimental error is too large to permit a differentiation between the heats of dissociation of light and heavy acetic acids.

Davies and Sutherland (1938) tried to find an experimental value of the heat of dissociation of acetic, benzoic and lauric acid in dilute solution. Unfortunately, their measurements did not extend over a sufficient number of concentrations to lead to conclusive results. They investigated the integral absorption of the monomeric OH band at 3 530 cm^{-1} and the corresponding C=O band at 1 770 cm^{-1} over a range of three different concentrations and temperatures assuming that the mass action law holds. If c is the concentration in mole per litre and α the concentration of monomers in the solute, then $c\alpha = (Kc/2)^{1/2}$ where K is the equilibrium constant for the equilibrium between monomers and dimers if $\alpha \ll 1$. Since α is proportional to m/c (m = integrated absorption), m^2/c should be a constant for different concentrations. The experimental values of m^2/c vary, however, by as much as 50%. The authors obtained a heat of association of $9{\cdot}3 \pm 1{\cdot}0$ kcal/mole for acetic acid and of 6·6 kcal/mole for benzoic acid, which latter was shown by Davies (1938) by the cryoscopic method to possess a heat of dissociation of 14·7 kcal/mole in solution. Davies and Sutherland ascribed the deviation of the spectroscopic results from the thermodynamical determinations to an effect of the concentration and temperature on the absorption coefficient. They confirmed this hypothesis by an investigation of the absorption of cetyl

alcohol as a function of concentration and temperature in highly dilute solutions where the association band was absent. Fox and Martin (1940), however, could show that the molecular absorption coefficient of the monomeric band of benzyl alcohol is quite independent of the concentration over a range of 0·0005 to 0·03 mole/litre and that the temperature effect is very slight.

Mecke and his collaborators made a careful and accurate study of the intensity of the second overtone of the monomeric O–H band for aromatic and aliphatic alcohols over a range of concentrations from 0·0375 mole/litre to 6 mole/litre. They used a spectrometer of higher resolving power so as to eliminate overlapping from neighbouring C–H bands and the O–H association band; the temperature of the solutions was kept constant by means of a thermostat. Care was taken to correct the observations for stray light, for errors in the setting of the cell and for a deviation of the cell windows from the parallel direction. The first investigation was carried out on the monomeric O–H band of a solution of phenol in CCl_4 at 10 320 cm^{-1} (Kempter and Mecke 1939, 1940). The band is very sharp and has a half-width of 100 cm^{-1} so that the integration over the band can be replaced by the maximum absorption coefficient ϵ_c which is determined as $(1/cd)\log I_0/I$ where I/I_0 is the minimum transmission obtained with a solution of concentration c mole/litre in a cell of length d cm. α, the degree of dissociation, i.e. the fraction of single phenol molecules, is then given by ϵ_c/ϵ_0 where ϵ_0, the absorption coefficient of the single molecule for finite dilution, is obtained by extrapolating the experimental values of ϵ_c to $c=0$. The authors measured the absorption coefficient for eight concentrations to an accuracy of within $\pm 2\%$ for a band of maximum absorption of 5%. α decreases with increasing concentration as the number of monomers is diminished at the expense of the polymerized molecules. Kempter and Mecke assumed at first that, over the whole range of concentrations investigated, the phenol molecules associate to dimers and that the mass action law holds. If the equilibrium constant is denoted by K, then $c\alpha^2 = K(1-\alpha)/2$, and a plot of $c\alpha^2$ against α should give a straight line. It was found that for concentrations above 0·03 the observed points deviate from linearity; for higher concentrations, the polymerization into larger complexes predominates over the bimolecular association. If it is now assumed that every complex of n molecules is formed by addition of a single molecule to the complex of degree of association $n-1$ and that the equilibrium constants are identical, then $c\alpha\alpha_n/n = K\alpha_{n+1}/(n+1)$ and

$$\alpha_n = n\alpha(\alpha c)^{n-1}/K^{n-1}. \qquad \ldots\ldots(1)$$

But since $\sum_1^\infty \alpha_n = 1$, $\qquad c\alpha = K(1-\sqrt{\alpha}). \qquad \ldots\ldots(2)$

The values obtained for $c\epsilon_c$ were therefore plotted against $\sqrt{\epsilon_c}$ and a straight line obtained. From the slope and intercept, the following values for ϵ_0 and K are determined: $\epsilon_0 = 0{\cdot}815 \times 10^{-4}$ cm^2/mole, $K = 0{\cdot}443 \pm 0{\cdot}005$ mole/litre. The concentration of an n-fold molecule α_n can then be determined from (1) for every concentration (table 8).

The intensity of the O–H fundamental of phenol at 3 620 cm^{-1} has been investigated by Fox and Martin (1937). Fox and Martin took the same precautions as Kempter and Mecke and purified their solvents by fractionation and drying. If their results for $c\epsilon_c$ are plotted against $\sqrt{\epsilon_c}$, a fairly straight line is obtained, yielding $K = 0{\cdot}49 \pm 0{\cdot}1$ and $\epsilon_0 = 200 \pm 10$ in the same units as before for a range

of concentrations from 0·0145 to 1·174 mole/litre. Values of $\alpha_1 = 86\%$ and 64% can be evaluated for concentrations of 0·0375 and 0·15 respectively, which check very well with the values for the second overtone (table 8).

Kreuzer and Mecke (1941) tried to apply the same treatment to the spectrum of the primary aliphatic alcohols after Kreuzer (1941) had shown that the absorption coefficient of the C–H vibration at 11 000 cm^{-1} (third overtone) is additive, i.e. proportional to the number of C–H valencies in the molecule. The plot of $c\epsilon_c$ against $\sqrt{\epsilon_c}$ results in a curved line for methyl, ethyl, propyl and butyl alcohol so that it must be concluded that there is an upper limit for the number of molecules in a multi-molecular complex. Furthermore, the equilibrium constants in the mass action law can no longer be assumed to be identical. The tangents to the curves of ethyl, propyl and butyl alcohol form

Table 8. Percentage Concentration of n-fold Molecule α_n as Function of n and of Total Concentration c

	Total concentration c (mole/litre)				
n	0·0375	0·15	0·30	1·5	3·0
1	86·1	62·5	47·0	17·1	10·1
2	12·2	26·3	29·6	19·9	13·8
3	1·3	8·3	14·0	17·5	14·1
4	0·28	2·3	5·9	13·7	12·8
5	0·12	0·61	2·3	10·0	10·9
>5	0·0	0·0	1·2	21·8	38·3

an angle of 90° with the abscissa for $c \to 0$. Since the first differential coefficient is inversely proportional to the equilibrium constant K_{12} ($c_1^2 = K_{12}\alpha_2/2$), K_{12} appears to be zero for these alcohols, i.e. there are no dimers present in the mixture of molecular aggregates. If it is assumed that $\alpha_2 = 0$ and all K_{ij} identical for n greater than 2, the experimental results can be satisfactorily represented by the constants $K = 0{\cdot}33$ mole/litre and $\epsilon_0 = 0{\cdot}055$ cm^2/mole for ethyl alcohol, $K = 0{\cdot}32$ and $\epsilon_0 = 0{\cdot}058$ for propyl alcohol and $K = 0{\cdot}30$ and $\epsilon_0 = 0{\cdot}063$ for butyl alcohol. The constants, therefore, are seen to increase with growing length of molecular chain.

Since these results differed so widely from the behaviour of phenol, Hoffmann (1943) investigated the effect of the water content of the solute and solvent on the intensity of the OH band by careful drying. Care was taken to use only freshly distilled solvent in the preparation of the solutions, and the solute was directly distilled into the measuring cylinder about to be weighed. It was found that the monomeric band remains unaffected by the water content of solute and solvent alike, a result which had already been obtained for the solvent by Fox and Martin (1940). It appears therefore that the water molecules attach themselves only to larger molecular complexes and have consequently an effect on the intensity of the association band. The measurements of the intensity of the OH band of phenol were found to confirm Kempter's and Mecke's results for higher concentrations although the plot of $(c\alpha, \sqrt{\alpha})$ shows a slight curvature for small concentrations. Similarly, the curved character of the lines representing $c\alpha$ (or $c\epsilon_c$) as function of $\sqrt{\alpha}$ (or $\sqrt{\epsilon_c}$) was unchanged by the careful purification of the substances. The investigations covered a range of concentrations from

very small values up to 2·5 mole/litre (for higher concentrations the mass action law cannot be expected to hold) and included a number of secondary and tertiary alkyl alcohols as well as cetyl alcohol as an example of a very long chain molecule ($C_{16}H_{33}OH$) and benzyl alcohol ($C_6H_5CH_2OH$) with its large benzene ring. Hoffmann assumes now that the number of single molecules n which can form an n-fold polymer is limited and that furthermore the equilibrium constants are all different. The solution consists then of a mixture of f-fold polymers of concentration α_f where f assumes all values between 1 and n. The individual concentrations α_f are determined by the mass action law, where every f-fold polymer is produced by a reaction between a single molecule and a $(f-1)$-fold polymer. The possibility of reactions between polymers themselves has been completely left out of account. Equation (1) is then modified to:

$$\alpha_n = n\alpha(\alpha c)^{n-1}/\prod_2^n K_{(i-1)i} \qquad \ldots\ldots(1a)$$

where $K_{(i-1)i}$ denotes the equilibrium constant of the reaction between a single molecule and an $(n-1)$-fold polymer. $\Sigma_1^n \alpha_i = 1$ but the summation cannot be carried out now. If $\alpha = \epsilon_c/\epsilon_0$ as before, a mean number $\bar{f}$ can be introduced in such a way that the total concentration c becomes

$$c = \bar{f} \int_0^{\epsilon_c} \frac{1}{\epsilon_c} d(\epsilon_c c) \qquad \ldots\ldots(3)$$

where the integral is obtained from the experimentally determined values of the absorption coefficients ϵ_c. $\bar{f}$ has then the meaning of an average degree of polymerization. The results show a grouping of the alcohols into primary, secondary and tertiary alcohols: the tendency to association depends therefore on the position taken up by the OH group in the molecule and is comparatively independent of the structure of the rest of the molecule. It follows furthermore from the curves that the formation of dimers is less likely than that of triply and fourfold polymerized molecules for all the alcohols under investigation, and it appears that the probability of association into even higher polymers decreases again.

The influence of the solvent on the absorption intensity has been studied for the case of t-butyric alcohol; CCl_4, CS_2 and cyclohexane were investigated. The results show that cyclohexane and CS_2 have a strong effect on the polymerization; $\bar{f}$ reaches a saturation value of three in both solvents for higher concentrations at room temperature (21·5° c). In this case, equilibrium is established between single and triple molecules, and the equilibrium constant of this reaction can be determined from the absorption coefficients. It has a value of $0{\cdot}0032 \pm 0{\cdot}0006$ mole^2/litre^2 in cyclohexane and $0{\cdot}061 \pm 0{\cdot}004$ mole^2/litre^2 in CS_2. The absorption coefficients of t-butyric alcohol were also investigated at two higher temperatures, 30° c and 40°c, and a mean heat of association of $5{\cdot}3 \pm 0{\cdot}5$ kcal per mole linkage obtained.

All the experiments mentioned above, with one exception, refer to the connection between the intensity of the monomeric band and the number of single molecules at different concentrations. Various experimenters (Fox and Martin 1937, 1940, Errera, Gaspart and Sack 1940) studied the intensity of the association band at different concentrations and temperatures and tried to determine the concentrations of the polymers responsible for this band. This problem is much more intricate since the broad association band is produced

by a mixture of different polymers whose individual concentrations and true absorption coefficients (called ϵ_0 in the case of the monomer) are unknown. Only when the association band consists of one kind of molecular aggregate—as in the case of the CH_3COOD dimer investigated by Herman and Hofstadter (1938)—will conclusive results be obtainable. Fox and Martin (1940) examined the intensities of the monomeric and dimeric bands of benzyl alcohol in the region of the O–H fundamental. Their observations covered a range of concentrations from 0·0005 to 0·174 mole/litre. At concentrations between 0·00664 and 0·0338 mole/litre, the association band is fairly narrow with a maximum at 3 500 cm^{-1}; at $c = 0{\cdot}067$ mole/litre, the band has broadened considerably on the short frequency side of the maximum. At $c = 0{\cdot}170$ mole/litre, the maximum of the association band has shifted to 3 340 cm^{-1} but there is a distinct shoulder in the 3 500 cm^{-1} region. Fox and Martin attribute this band to the dimers in the solution and show that the maximum absorption coefficient ϵ_{2c} of the dimer band divided by the concentration c is a constant; it varies between 154 for $c = 0{\cdot}00664$ and 150 for $c = 0{\cdot}0338$. This result would mean that for small concentrations there are only monomers and dimers present with the dimer concentration small compared with that of the monomers in direct contradiction to Hoffmann's conclusions. These authors concluded furthermore that the measurements of the intensities of the dimeric and polymeric bands are compatible with an equilibrium between double and quadruple molecules with an equilibrium constant of 84. These conclusions are however invalidated by the fact that the true absorption coefficients of single, double and triple molecules are assumed to be identical.

The heat of transition from the *trans*-configuration to the *cis*-configuration of *o*-chlorophenol has been determined by Davies (1938). He measured the integrated absorption of the two components at 3 600 and 3 540 cm^{-1} at 73°C and assumed their intensity ratios 1/7·98 to be equal to $\exp(-E/\boldsymbol{R}T)$, where E is the energy of the transition, $\boldsymbol{R}$ the gas constant and T the absolute temperature, resulting in $E = 1{\cdot}4$ kcal/mole. Pauling (1936), who was the first to put forward the interpretation of a *cis–trans* isomerism for the splitting of the first overtone of the O–H band, had obtained the same value at room temperature. However, as Lüttke and Mecke (1950) have pointed out, these results are really inconclusive since the true absorption coefficients of the two forms may not be identical as has been assumed in these calculations. The heat of association of O–H groups has been deduced by Fox and Martin (1937) in a completely different manner. They obtained the heat of dissociation D of the O–H linkage from the anharmonicity constant x to $D = \boldsymbol{h}\nu_0^2(1-x)/4x\nu_0$ where ν_0 is the natural frequency of the O–H vibration for small displacements and the nth overtone $\nu_n = n\nu_0\{1-(n+1)x\}$. If x remains the same for the monomeric band and association band in solution, then $\delta D/D = \delta\nu_0/\nu_0$ where δD is the difference in the dissociation energies of the unassociated and associated linkage and $\delta\nu_0$ the corresponding frequency change. Since D has a mean value of 110 kcal for the unassociated alcohols and the association band lies 150–300 cm^{-1} below the monomeric band, δD will lie between 3 and 6 kcal/mole.

In conclusion it may be pointed out that the heat of association appears to be higher for carboxylic acids than for alcohols, confirming the greater tendency to association of the acids.

§5. THE HYDROGEN BOND. DISCUSSION OF THE FORCES RESPONSIBLE FOR THE ASSOCIATION BAND

It has been shown that the vibration of the hydroxyl group undergoes a relative displacement of the order of 10% (table 6) when the hydroxyl compound makes the transition from the gas to the liquid phase; the O–H binding force is therefore lowered by about 20% compared with that of the gas. Cryoscopic measurements have produced evidence that in the liquid the molecular weight is increased, i.e. the single molecules of the gas have associated into larger aggregates held together by forces which are of the order of 5–8 kcal/mole as is known from the investigation of the heat of association. This association energy amounts to approximately 7% of the binding energy of the O–H bond. In a dilute solution, two bands appear; one, the monomeric band, is in the position of the vibration of the single molecule, and the second, of lower frequency, is ascribed to the O–H vibration of the associated polymer. As the concentration increases, the monomeric band decreases in intensity without changing its width appreciably while the association band grows in width and intensity with a corresponding lowering of the frequency of the absorption maximum. There is no trace of the monomeric band in the spectrum of the pure liquids. With a closer approach of the molecules, the type of association therefore seems to alter. In the case of phenol, it appears to be proved that the degree of polymerization increases with increasing concentration and that in the liquid very long chains of polymers are formed; the polymers of various lengths co-exist in solutions of high concentration. There exists an attractive force between the hydrogen atom chemically linked to a molecule (usually attached to a hydroxyl group) and an oxygen atom—or S, Cl, N and F—of a different (intermolecular association) or the same molecule (intramolecular association). This molecular interaction force comes into play when the approach between the interacting groups is sufficiently close; it is of approximately the same strength whatever the structure of the substances involved although it is definitely larger for carboxylic acids than for alcohols. As a result, a special type of bonding is produced much weaker than a valency bond, which leads to the formation of polymers and to which the term ' hydrogen bond ' is now generally applied. As soon as a hydrogen bond is effected, the O–H valency linkage of the hydroxyl group taking part in the association is loosened; the hydrogen is pulled away from the molecule to which it is chemically linked and attracted towards the associated group. The mechanism by which a hydrogen bond is produced is still very imperfectly understood. Various explanations have been put forward which will be discussed in this section.

Bauer and Magat (1938) made an attempt to account for the frequency shift of the O–H vibration when the water molecule passes from the gaseous to the liquid and solid state by purely electrostatic forces. They use the Bernal and Fowler (1933) model for the structure of liquid water: every oxygen atom is tetrahedrally surrounded by four others at a distance of 2·76 Å at room temperature. This tetrahedral unit persists into the liquid state. The tendency of the oxygen atoms to form clusters with an O–O distance of 2·6–2·8 Å has been confirmed by an analysis of the x-ray spectra of liquid alcohols and carboxylic acids (Pierce and McMillan 1938). Kreuzer and Mecke (1941) explain the absence of dimers in alcohols by the fact that the tetrahedral coordination is first realized in triple

molecules. Bauer and Magat consider the electrostatic forces between one water molecule and the four nearest neighbours where every nucleus is assumed to carry an effective charge. They obtain a relationship between the frequency shift and the distance between neighbouring atoms which latter is known from the Bernal–Fowler model. As the result of their calculations they find a frequency displacement of half of the observed value; they try furthermore to account for the width of the association band by the variation of the interaction forces due to the variation in the distances of the molecules. Their theory has a number of serious flaws: they completely neglect the polarization of the central molecule by its neighbours and the reaction of the centre—effects which partially cancel the initial interaction forces and therefore reduce the band displacement even further. The magnitude of the effective charge is largely conjectural and is chosen *ad hoc* to give the right order of magnitude. As Sutherland (1940) has pointed out, there is another serious objection to the idea of attributing sole responsibility for the energy of association to the electrostatic interaction: if this were true, all substances with an X–H bond possessing a large dipole moment should show similar association. But, for instance, H_2S, NH_3 and PH_3 show sharp bands in the liquid state and a much smaller displacement of the frequency from the position in the gas.

Fox and Martin (1937) also discuss the relationship between the van der Waals forces and the energy of association of hydroxy compounds. They find that a heat of association of 5·4 kcal/mole is obtained if the dipole–dipole interaction, induction effect and dispersion forces are taken into account and the three forces are assumed to be in the ratio 190 : 10 : 47 if the O–H dipole moment has the value $1{\cdot}7 \times 10^{-18}$ e.s.u. This heat of association is of the right order as comparison with the experimental values shows (§ 4). However, the van der Waals potential energy does not vary sufficiently with distance to produce the required frequency shift on association. Fox and Martin (1940) derived a semi-empirical formula relating the heat of dissociation D to the force constant k of the bond: $D \propto k^{3/5}$ which leads to the equation $dk/k = \frac{5}{3} dD/D$. For water dD the heat of association is obtained from the heat of sublimation of ice to 5·8 kcal per bond and $D = 110$ kcal. It follows that the change in the force constant dk on the transition from vapour to ice becomes $-0{\cdot}088k$ corresponding to a frequency change from 3 700 cm^{-1} to 3 540 cm^{-1} as compared with the experimental value of 3 259 cm^{-1}. The experimental investigation of the influence of different solvents on the O–H frequency of phenol (Lüttke and Mecke 1949) has also shown that the dipole–dipole attraction is of relatively small importance for the formation of a hydrogen bond.

It must therefore be assumed that the mechanism of hydrogen bond formation bears some resemblance to that of a covalent bond. Mecke (1944) expresses the idea that two steps are necessary to produce a hydrogen bond involving both van der Waals and exchange forces. The van der Waals forces loosen the O–H valency bond; as a result the proton acquires—possibly as a consequence of the superposition of electronic charge clouds—quantum-mechanical bonding tendencies to two partners, one of which is oxygen, but without taking up a symmetrical position between the two. In fact, it is still attached by a mainly covalent bond to the oxygen of the hydroxyl group with a slightly reduced binding strength. Several attempts have been made to calculate the length of the O–H valency bond in associated molecules from the values of the force constants for the monomeric and polymeric bands (Cross, Burnham and Leighton 1937,

Fox and Martin 1940, Sutherland 1940). The results give a shift in the equilibrium distance of 0·016 to 0·04 Å from the O–H distance of 0·97 Å in the free molecule for water and alcohols and 0·09 Å in carboxylic acids. Consequently, the hydrogen bond distance cannot be less than 1·7 Å. Lüttke and Mecke (1949) came to the conclusion in their study of the influence of the solvent on molecular association that a hydrogen bond is formed if a hydroxyl group with an active hydrogen atom approaches a group with easily polarizable π-electrons. The presence of π-electrons in the donor is therefore a necessary but not sufficient condition for hydrogen bond formation. In alcohols and water, every hydroxyl group acts as acceptor and donor simultaneously; the π-electrons are provided by the oxygen. In carboxylic acids, the hydroxyl group is the acceptor and the C=O bond delivers the necessary charge to maintain the hydrogen bond, resulting in a redistribution of the electronic cloud responsible for the double bond and consequently in a weakening of this bond and corresponding lowering of the C=O frequency (table 3).

The intramolecular association is interpreted by Lüttke and Mecke as the result of the mesomeric and inductive effects for which Hückel gave the following definition: The mesomeric effect can occur for all substituents (in the benzene ring) which possess π-electrons or are capable of accepting them. Substituents with unlinked electron pairs can deliver π-electrons; they can be accepted by substituents with a double bond capable of conversion into two single bonds. The mesomeric effect is of noticeable importance only for substituents in the *ortho* and *para* position. With a substituent which can deliver π-electrons, the charge density of π-electrons in the benzene ring is increased in the *o* and *p* position; it decreases with a substituent of the second kind. The polarity of the bond between the carbon atom of the ring and the substituent (π-electron bond) produces a perturbation in the charge distribution of the ring; this perturbation reaches a maximum for the *o* and *p* position, producing an alternating increase and decrease of the density of the π-electrons. This effect is called the inductive effect.

In general, the mesomeric and inductive effects of any substituent oppose each other; they can therefore only have an influence if they are different in magnitude. The mechanism by which intramolecular association comes about can then be explained in the following manner for the example of *o*-chlorophenol: the π-electrons delivered by the oxygen of the hydroxyl group to the benzene ring flow through the ring into the C–Cl linkage producing an attraction between the C–Cl and O–H linkages. This process is only possible if the hydroxyl shows a larger mesomeric and smaller inductive effect than the halogen. The interaction between chlorine and the O–H group is however small in consequence of the fact that the mesomeric as well as the inductive effects have the same direction in the two substituents, so that an equilibrium between the *cis* and *trans* configurations is established as a result of the thermal agitation of the molecules. This interpretation is applicable to all the substituents of the first class investigated by Lüttke and Mecke (1950) mentioned in § 3. In the compounds of the second class, the two effects have the same direction for the two substituents; in both groups the mesomeric effect will have to predominate over the inductive effect if intramolecular interaction is to take place. If this condition is fulfilled, π-electrons are delivered by the O–H group to the benzene ring, drawn away again by the substituent of the second class (for instance, a C=O group) and thence

attracted by the proton of the hydroxyl group leading to an overlapping with the electronic cloud of the O–H bond. A closed ring-shaped charge distribution is thus formed between the two substituents; in this case the hydrogen bond has more the character of a covalent link than in a case of the first class. The infra-red intensity of the O–H band is therefore less in compounds of the second class. The great width of the band is tentatively attributed to a combination of the O–H vibration with low rotation frequencies of the O–H and donor groups. It would presumably mean that the position in space of the two valency links taking part in the interaction is not fixed but varies slowly all the time.

This wave-mechanical interpretation of the hydrogen bond is of necessity only qualitative. It shows that the binding force must be of the second order compared with the O–H valence force. Direct confirmation of the existence of a hydrogen bond would be obtained if it were possible to determine experimentally the frequency of its vibration which must be expected to be very low. It is very difficult to estimate its value since we only know how much the O–H valency force is weakened by association but nothing of the actual hydrogen bond strength. Fox and Martin (1940) have tried to compute the energy of the hydrogen bond D_2 in the following manner: if D_0 is the energy of dissociation of an unassociated O–H link and this energy is changed to D_1 on association, then $D_1+D_2-D_0$ will be the energy of association determined experimentally. By combining a modified Morse–Clark relation of the form $kr^5=$constant with an equation derived by Sutherland (1938) of the form $D/kr^2=$constant, the equation $Dr^3=$constant is obtained where k is the force constant of the vibration, r the bond length at equilibrium and D the dissociation energy of the bond. If this equation is applied remembering that D_0, r_0 and $D_1+D_2-D_0$ as well as r_1+r_2 (O–O distance in the state of association) are known, the following values for D_1, D_2, r_1 and r_2 are obtained: $D_1=96{\cdot}3$ kcal/mole; $D_2=19{\cdot}5$ kcal; $r_1=1{\cdot}014$ Å; $r_2=1{\cdot}726$ Å if $D_0=110$ kcal, $r_0=0{\cdot}970$ Å, $D_1+D_2-D_0=5{\cdot}8$ kcal and $r_1+r_2=2{\cdot}74$ Å (values for ice at 0° K). D_2 is then the energy of dissociation of the hydrogen bond and is seen to be surprisingly high. If f_0 the force constant of a free O–H link is obtained from the O–H vibration 3 700 cm^{-1} as $7{\cdot}54\times10^5$ dyn/cm, f_1 the force constant of the associated O–H bond and F the force constant of the hydrogen bond can then be calculated from one of the above equations and are found to have the values: $f_1=5{\cdot}99\times10^5$ and $F=0{\cdot}42\times10^5$ dyn/cm. If the O–H valency linkage and the hydrogen bond are assumed to be collinear, the vibrations along the bond axis using the above values of f_1 and F can be computed to lie at 3 400 and 284 cm^{-1} and at 2 470 and 186 cm^{-1} for heavy hydrogen. The value of 3 400 cm^{-1} is much too high as compared with the experimental value (table 2). On the other hand, an infra-red frequency of 160 cm^{-1} and a Raman frequency of 200 cm^{-1} have been observed in the spectra of ice and light and heavy water. Sutherland (1940) had already tentatively identified this frequency with the hydrogen bond oscillation. In any case, the above calculations can only give an idea of the approximate magnitude of f, F and D_2 since they are based on the assumption that the potential function of the hydrogen bond has the same form as that of the valency link.

So far, all information about the hydrogen bond has come from the study of the O–H valency vibration of monomers and associated molecules. H_2O in the associated state shows another band in an easily accessible region of the spectrum at 2 100–2 200 cm^{-1} which seems to owe its existence to the intermolecular forces

since it shifts from higher frequencies in ice to lower frequencies when the temperature is raised and loses at the same time in intensity. It has been interpreted by Ellis (1931) as a summation tone of the deformation frequency 1 640 cm^{-1} and a frequency of hindered rotation of the H_2O lattice near 600 cm^{-1} (centre of the broad band 320–1 020 cm^{-1}). This explanation seems very doubtful in view of the fact that the 2 150 cm^{-1} band has been observed in mixtures of water and dioxane where it persists down to concentrations of 13% (by volume) of water (Gordy 1936) in which the lattice vibration 600 cm^{-1} cannot occur since it is generally supposed that in a dioxane–water mixture the hydrogen bonds form between the hydrogen atoms of the water and the oxygen atoms of the dioxane. Such a polymer cannot be expected to possess the same lattice vibration as water; besides, it seems very unlikely that the lattice-like structure will be sufficiently established at lower concentrations. It would therefore be of interest to make a further investigation of this band in other solvents, oxygenated as well as non-polar, of varying concentrations.

A characteristic property of the association bands is their extreme broadness which is usually, somewhat vaguely, attributed to the superposition of very slow torsional frequencies over the valency vibration. Now, if the associated molecular complex is assumed to form a long chain polymer where the molecules are coupled together through their O–H bonds forming a hydrogen bond with adjacent molecules, a spectrum of O–H frequencies must be expected. In order to obtain an estimate of the width of such a spectrum, the vibrations of a polymer of the type $(ROH)_n$ have been calculated. n molecules ROH are supposed to be arranged in a chain in such a manner that the O–H linkages lie in one plane and are linked together by hydrogen bonds while the radicals R will lie outside this plane. The oxygens have a distance of 2·76 Å and form a zigzag chain with a tetrahedral angle in conformity with the x-ray data for alcohols and water. The hydrogen bond is usually assumed to be collinear with the O–H valency bond although there is no direct evidence for this. Such a co-planar chain, whose recurrent structure will consist of two oxygen and two hydrogen atoms, will show a spectrum of four branches: one branch is due to the vibrations out of the plane of the chain involving the bending of the O–H · · · O line where the dots indicate the hydrogen bond; the second branch is in the main produced by the deformation of the tetrahedral angle, the third by the hydrogen bond stretching and the fourth by the O–H valence oscillation. The calculations have been carried out for an infinitely long chain; all the frequencies of the polymers $(ROH)_n$ of all possible degrees of polymerization n will then lie within the range of these four branches which therefore give the greatest possible widths of bands to be expected. The first branch, by the way, is quite independent of the forces responsible for the other three. The following symbols have been chosen for the force constants: f = O–H valence force constant, F = hydrogen bond force constant, H = angle force constant producing a change in the tetrahedral angle with an oxygen at the apex, G = force constant producing the bending of the chain out of its plane. The influence of the radicals has been neglected since they can be assumed to be too heavy to contribute more than a second order effect to the O–H vibrations. The values obtained for the O–H valency frequencies are given in table 9 for various force constants F and H. The O–H valence force constant f has been taken to be 6×10^5 dyn/cm. The first column of table 9 contains H, the angle force constant; the limiting frequencies of the

O–H vibration branch are given for hydrogen bond force constants F of the value 0·1, 0·3 and $0{\cdot}6 \times 10^5$ dyn/cm. It will be seen that the width of the band depends only on the magnitude of the angle force H and is completely independent of the strength of the hydrogen bond. The position of the band, on the other hand, is mainly determined by F and f. This explanation of the width of the

Table 9. The O–H valence vibration branch of an infinitely long zigzag chain $(\text{O–H} \cdots \text{O})_n$ of alternating O–H valence bonds and hydrogen bonds

	Frequency range (in cm^{-1}) of O–H valence vibrations for		
H (dyn/cm)	$F = 0{\cdot}1 \times 10^5$ dyn/cm	$F = 0{\cdot}3 \times 10^5$ dyn/cm	$F = 0{\cdot}6 \times 10^5$ dyn/cm
$0{\cdot}06 \times 10^5$	3330–3390	3380–3440	3440–3500
$0{\cdot}18 \times 10^5$	3330–3490	3380–3540	3450–3610
$0{\cdot}30 \times 10^5$	3340–3610	3390–3660	3460–3730

association band is applicable to alcohols and water but fails in the case of the carboxylic acids where the association produces dimers and not long chain polymers. However, the association band of the acids is extremely broad as has been mentioned in §2. Since Pierce and McMillan (1938) found an x-ray pattern for acids similar to that of the alcohols, they concluded that there must be an orderly arrangement of the oxygen atoms, perhaps in such a manner that the O–H $\cdots$ O links lie parallel to each other preserving the tetrahedral coordination of the oxygen atoms. In that case, angle forces of the type of the constant H may come into play coupling the dimers together and causing thus the broadening of the O–H valence vibration.

In conclusion, it may be said that the hydrogen bond is formed between a hydroxyl or amino group and an easily polarizable group of the same or a different molecule. It leads to intermolecular and intramolecular association and can be stable enough to persist into the gaseous state. It is of much lower strength than a valence bond although too large to be accounted for by van der Waals forces alone and may therefore be assumed to be due to a partial overlapping of the wave functions of the groups concerned. A hydrogen bond can only occur between hydrogen and an atom possessing π-electron pairs which do not take part in a valency bond. The effect is consequently entirely confined to the elements fluorine, oxygen, nitrogen, and carbon only if one of its valence links is double-bonded.

REFERENCES

Badger, R. S., and Bauer, S. H., 1937, *J. Chem. Phys.*, **5**, 839; 1940, *Ibid.*, **8**, 288.
Bauer, S. H., and Magat, M., 1938, *J. Phys. Radium*, **9**, 319.
Bernal, J. D., 1937, *Trans. Faraday Soc.*, **33**, 142, 210.
Bernal, J. D., and Fowler, R. H., 1933, *J. Chem. Phys.*, **1**, 515.
Bloch, M., and Errera, J., 1935, *J. Phys. Radium*, **6**, 154.
Borst, L. B., Buswell, A. M., and Rodebush, W. H., 1938, *J. Chem. Phys.*, **6**, 61.
Bosschieter, G., 1937, *J. Chem. Phys.*, **5**, 992.
Bosschieter, G., and Errera, J., 1937, *J. Phys. Radium*, **8**, 229.
Bosschieter, G., Errera, J., and Gaspart, R., 1938, *Physica*, **5**, 115.
Brattain, R. R., 1938, *J. Chem. Phys.*, **6**, 298.
Buswell, A. M., Deitz, V., and Rodebush, W. H., 1937, *J. Chem. Phys.*, **5**, 84, 501.
Buswell, A. M., Maycock, L. S., and Rodebush, W. H., 1940, *J. Chem. Phys.*, **8**, 362.
Buswell, A. M., Rodebush, W. H., and Roy, M. F., 1938, *J. Amer. Chem. Soc.*, **60**, 2239, 2528.
Cartwright, E., 1936, *Phys. Rev.*, **49**, 470.
Cartwright, E., and Errera, J., 1936, *Proc. Roy. Soc.* A, **154**, 138.

COGGESHALL, N. D., 1947, *J. Amer. Chem. Soc.*, **69**, 1620.
CROSS, P. C., BURNHAM, J., and LEIGHTON, P. A., 1937, *J. Amer. Chem. Soc.*, **59**, 1134.
DAVIES, M. M., 1938, *Trans. Faraday Soc.*, **34**, 410, 1427; 1940, *Ibid.*, **36**, 333; 1948, *J. Chem. Phys.*, **16**, 267.
DAVIES, M. M., and SUTHERLAND, G. B. B. M., 1938, *J. Chem. Phys.*, **6**, 755, 767.
ELLIS, J. W., 1931, *Phys. Rev.*, **38**, 693.
ERRERA, J., 1937, *J. Chim. Phys.*, **34**, 617.
ERRERA, J., GASPART, R., and SACK, H., 1940, *J. Chem. Phys.*, **8**, 63.
ERRERA, J., and MOLLET, P., 1935, *C.R. Acad. Sci., Paris*, **200**, 814, *J. Phys. Radium*, **6**, 281; 1936, *Nature, Lond.*, **138**, 882; 1937, *C.R. Acad. Sci., Paris*, **204**, 259.
ERRERA, J., and SACK, H., 1938, *Trans. Faraday Soc.*, **34**, 206.
FOX, J. J., and MARTIN, A. E., 1937, *Proc. Roy. Soc.* A, **162**, 419; 1938, *Ibid.*, **167**, 257; 1939, *Nature, Lond.*, **143**, 199; 1940, *Proc. Roy. Soc.* A, **174**, 234, *Trans. Faraday Soc.*, **36**, 897.
FREYMAN, R., 1933, *Ann. Phys., Paris*, **20**, 243; 1936, *J. Phys. Radium*, **7**, 476, 506; 1937, *C.R. Acad. Sci., Paris*, **204**, 41.
FRIEDEL, R. A., 1951, *J. Amer. Chem. Soc.*, **73**, 288.
GILLETTE, R. H., and DANIELS, F., 1936, *J. Amer. Chem. Soc.*, **58**, 1139.
GILLETTE, R. H., and SHERMAN, A., 1936, *J. Amer. Chem. Soc.*, **58**, 1135.
GORDY, W., 1936, *J. Chem. Phys.*, **4**, 769; 1937, *J. Amer. Chem. Soc.*, **59**, 464, *J. Chem. Phys.*, **7**, 163; 1938, *J. Amer. Chem. Soc.*, **60**, 605; 1940, *J. Chem. Phys.*, **8**, 516.
GORDY, W., and NIELSEN, A. H., 1937, *J. Chem. Phys.*, **6**, 12.
GORDY, W., and STANFORD, S. C., 1940, *J. Amer. Chem. Soc.*, **62**, 497, *J. Chem. Phys.*, **8**, 170.
HERMAN, R. C., and HOFSTADTER, R., 1938, *J. Chem. Phys.*, **6**, 110, 534; 1939, *Ibid.*, **7**, 460, 802; 1940, *Ibid.*, **8**, 252.
HIBBEN, J. H., 1937, *J. Chem. Phys.*, **5**, 166.
HOFFMAN, E. G., 1943, *Z. phys. Chem.* B, **53**, 179.
HOFSTADTER, R., 1938, *J. Chem. Phys.*, **6**, 540.
HUGGINS, M. L., 1936, *J. Organic Chem.*, **1**, 407.
KEMPTER, H., and MECKE, R., 1939, *Naturwissenschaften*, **27**, 583; 1940, *Z. phys. Chem.* B, **46**, 229.
KINSEY, W., and ELLIS, J., 1936, *Phys. Rev.*, **49**, 105.
KREUZER, J., 1941, *Z. Phys.*, **118**, 325.
KREUZER, J., and MECKE, R., 1941, *Z. phys. Chem.* B, **49**, 309.
LASSETTRE, E. N., and DICKINSON, R. M., 1939, *J. Amer. Chem. Soc.*, **61**, 54.
LASSOTTRE, N., 1937, *Chem. Rev.*, **20**, 259.
LÜTTKE, W., 1950, *Chem. Ber.*, **83**, 571.
LÜTTKE, W., and MECKE, R., 1949, *Z. Elektrochem.*, **53**, 241; 1950, *Z. phys. Chem.*, **196**, 56.
LYON, W., and KINSEY, E. L., 1942, *Phys. Rev.*, **61**, 482.
MAGAT, M., 1934, *J. Phys. Radium*, **7**, 347.
MCDOUGALL, F. H., 1936, *J. Amer. Chem. Soc.*, **58**, 2585.
MECKE, R., 1944, *Z. Elektrochem.*, **50**, 57.
PAULING, L., 1936, *J. Amer. Chem. Soc.*, **58**, 94.
PIERCE, W. C., and MCMILLAN, D. P., 1938, *J. Amer. Chem. Soc.*, **60**, 779.
PLYLER, E. K., and WILLIAMS, S., 1936, *J. Chem. Phys.*, **4**, 157.
SUTHERLAND, G. B. B. M., 1933, *Proc. Roy. Soc.* A, **141**, 535; 1938, *Proc. Indian Acad. Sci., Bangalore* A, **8**, 341; 1940, *Trans. Faraday Soc.*, **36**, 889.
WEST, W., 1939, *J. Chem. Phys.*, **7**, 795.
WILLIAMS, D., GATICA, T., and GORDY, W., 1937, *J. Phys. Chem.*, **41**, 645.
WILLIAMS, D., HOFSTADTER, R., and HERMAN, R. C., 1939, *J. Chem. Phys.*, **7**, 802.
WULF, O. R., 1936, *J. Amer. Chem. Soc.*, **58**, 2290.
WULF, O. R., and LIDDELL, V., 1935, *J. Amer. Chem. Soc.*, **57**, 1464.
WULF, O. R., LIDDELL, V., and HENDRICKS, S. B., 1936, *J. Amer. Chem. Soc.*, **58**, 2287.

MASS SPECTROMETRY

By K. I. MAYNE
The Clarendon Laboratory, Oxford

CONTENTS

	PAGE
§ 1. Introduction	24
§ 2. Design and instruction	25
2.1. General features outlined	25
2.2. First-order directional focusing	27
2.3. Second-order directional focusing	27
2.4. Stray field effects	30
2.5. Some practical arrangements	31
2.6. Two-directional focusing	32
2.7. The trochotron	33
2.8. Ion beam detection and measurement	33
2.9. Ion sources	37
§ 3. Nuclear measurements	39
3.1. Scope of nuclear measurements	39
3.2. Fission yields	39
3.3. Half-life determinations	41
3.4. Slow neutron absorption cross sections	42
3.5. Chemical atomic weights; some new rare isotopes	42
3.6. Branching ratios	43
§ 4. Isotopic abundance variations in nature	44
4.1. Thermodynamic properties	44
4.2. A paleo temperature scale	45
4.3. Isotopic abundance variations in nature	46
4.4. Nuclear effects	46
References	47

ABSTRACT. This report describes the recent progress in mass spectrometer design and applications. Improvements in the conventional single-focusing mass spectrometer, which have made possible the quantitative analysis of very small samples (as small as 10^{-15} g in some cases) with high precision, have been mainly in respect of increased resolving power by suitably shaping the magnetic deflecting field boundary (either symmetrically or asymmetrically) in order to obtain a significant third-order focusing effect in the source aperture angle, and of a greatly increased signal–noise ratio in the ion beam detector, by employing new devices, viz. the vibrating reed electrometer and the secondary electron multiplier. Double-focusing mass spectrographs are not discussed.

The sensitivity of the mass spectrometer is now such that small changes in isotopic abundance, involving as few as 10^7 nuclei in some cases, can be detected. Nuclear processes (neutron absorption, radioactive decay, branching, fission) producing a significant isotopic composition change have been studied, employing specialist techniques. The very small fractionations, due to the differing thermodynamical properties of the isotopes, which can now be measured have important geochemical implications; by this means also a paleo-temperature scale has been established.

§ 1. INTRODUCTION

In recent years, as instruments reliable and semi-automatic in operation have become available commercially, mass spectrometry has become a method of routine analysis in many application of physics, chemistry and biology (Boivin 1949, Neuert 1949, Eltenton 1949). This aspect of the subject will not, however, be dealt with here; but rather this report will describe how recent

improvements in instrument design and technique have enabled important problems of fundamental interest in physics and chemistry to be attacked by means of precise isotope abundance measurements. Instrumental improvements have been in the direction of increased resolving power and, more particularly, of greater sensitivity and precision of measurement of the ion beam currents.

The conventional homogeneous sector field geometry (a cylindrical lens arrangement) as adopted by Nier and others has a resolving power of the order of 100, which for most problems is sufficient; however, in many cases an improvement of this by a factor of 10 is to be desired, as, for example, in dealing with heavy molecules. Again, in those cases where an impurity or background peak of comparable intensity occurs isobaric with an isotope under investigation, unless the latter is resolved precise measurements of small changes in its abundance are impossible; such, for example, is the problem involved in the measurement of the abundance of ^{3}He, which has an isobar in HD, with a mass difference of approximately 1 part in 500.

The resolving power of the Nier geometry is limited largely by the aberrations of the aperture angle of the source, in which there is a first-order focusing effect. An improvement in the resolving power to the required extent, without the additional complication of a double focusing instrument of the Mattauch type, has been achieved in extending the directional focusing property of the sector shaped field to the second order by shaping the field boundaries suitably, having regard to the limitations imposed by the stray fields which occur there. It also appears possible to have a focusing in two directions (i.e. a spherical lens arrangement), at the same focal point, by means of an inhomogeneous sector field. The large solid angle of source which such an arrangement will accept makes it useful where beam intensity is very low.

At the same time, the sensitivity of detection of the ion beam current and the precision of comparison of two ion currents collected simultaneously has been considerably improved, so that changes in the relative isotopic abundance involving as few as 10^7 atoms are detectable.

These improvements have enabled the mass spectrometer to be applied to those problems in nuclear physics which result in a change of the isotopic composition (absorption and decay), and to those chemical thermodynamical processes which provide an isotopic fractionation. Many such processes occurred in nature throughout geological time, and the interpretation of the measured separations is of considerable interest to geologists and biologists.

Other improvements and techniques have been introduced in the field of mass spectrography, which is concerned with the accurate measurement of mass, by means of double focusing instruments (Mattauch and Herzog 1934, Duckworth 1950) or frequency measuring devices (Goudsmit 1948, 1949, Hipple and Thomas 1949, Hipple, Sommer and Thomas 1949); these, however, will not be dealt with in the article.

§2. DESIGN AND INSTRUMENTATION

2.1. *General Features outlined*

The mass selective property of a homogeneous magnetic field H, dispersing a beam of singly-charged positive ions of mass M, incident normally to it, according to the relation

$$\rho=(C/H)(2MV/e)^{1/2} \qquad \ldots\ldots(1)$$

(ρ being the radius of curvature of the ion path, V the energy of the ions in volts), continues to be the basis of instruments for the quantitative measurement of isotopic relative abundances. A particular geometry, but not the only possible one (Cartan 1934) which has proved convenient in practice, is that of the symmetrical sector-shaped homogeneous field (fig. 1, which is drawn in the medial plane), where an approximately collimated ribbon-shaped ion beam, produced in the source at A, enters and leaves the field perpendicular to the field boundaries OM, ON. This magnetic lens produces a directional focusing of the first order in the (small) aperture angle α of the ion beam at the exit slit B (Barber 1933, Mattauch and Herzog 1934). Values of the angle ϕ of 60° (Nier 1947) and 90° (Thode and Graham 1947) are particular cases, with some restrictions, of the general theorem which includes the value of 180° (semicircular focusing).

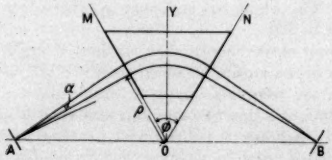


Fig. 1. Directional focusing, normal incidence symmetrical.

The resolving power $M/\Delta M$ of such an arrangement is largely determined by the finite widths of the source and exit slits, the aperture aberrations of the second order, and the energy spread of the ions in the beam; it can conveniently be made of the order of 100 with a magnet of a few hundreds watts power. At the same time the ion source A and ion-detecting device B are away from the disturbing influence of the deflecting field. With the more recent types of ion detector, in particular the electron multiplier, this is a convenient feature, while the ion source can be screened from stray magnetic fields, and so made less discriminating in mass in itself, as the deflecting field is tuned.

An attempt to increase the resolution of the sector arrangement, however, without increasing the magnet power requires a reconsideration of those factors producing aberrations at B of the object slit, particularly in regard to the shape of the field boundary (Hintenberger 1948, 1949, Kerwin 1949, 1950) and the influence of the stray fields (Ploch and Walcher 1950) which occur as the field bulges out over the pole edges in limiting the resolution realizable in practice.

Other devices, which possess mass selective properties, depending on the time of flight of the ions, have been described, as for example that of the ion drift tube, or the radio-frequency spectrometer (Bennett 1950). These, however, have the same limitations as regards the source of ions and the ion beam detector, limitations which are usually the most severe. In addition they have inherent disadvantages of resolution, so that they are unlikely to supplant the conventional deflection method for quantitative work, except in circumstances where their special properties of small dimensions and weight are of more importance than high accuracy in abundance measurements. The radio-frequency spectrometer has found use in rocket flights into the stratosphere and in work on negative ions which are too short-lived to travel the length of a Nier tube.

2.2. *First-order Directional Focusing*

Figure 1 illustrates the directional focusing property of an idealized symmetrical sector-shaped homogeneous magnetic field (i.e. one which is sharply defined by MON) in the medial plane. Ions produced in the source, and accelerated through a voltage V, emerge from the entrance slit of width s at A with a divergence α. Since the dispersion relation (eqn. (1)) contains V, the spread of energy of the ions in the beam will be dispersed in the magnetic field, contributing to the 'line' width at B. In some forms of double-focusing instrument this energy dispersion is compensated by an equal energy dispersion in an opposite sense in a preceding electric field (Mattauch and Herzog 1934), or is restricted by employing a preceding ion velocity filter. In single-focusing instruments it places a restriction on the type of ion source which can be used, and on the stability and ripple content of the accelerating voltage V, to such an extent that in the expression for the total line width d at B, viz.

$$d/\rho = \gamma + \delta + (\sigma + \alpha^2) + \text{higher terms in } \alpha \qquad \ldots\ldots(2)$$

(where $M = M_0(1+\gamma)$, $E = E_0(1+\delta)$, $\sigma = \rho s$, $\gamma = 0$ for one particular mass number) the term in δ is not predominant. A small contribution to δ due to a component v of the ion velocity in the direction of the field (z direction) can be neglected. Ion sources and high-voltage (~ 3000 v) power supplies (Nier 1947, Thode 1947, Urey *et al.* 1950) have been designed to produce ion beams isoenergetic to the extent that the energy spread is no significant restriction on single focusing, for a resolution of several hundreds.

The mass dispersion is given by γ in eqn. (2) for a mass difference $\Delta M = 1$, i.e.

$$D = \rho/M \text{ cm.} \qquad \ldots\ldots(3)$$

When this separation is just equal to the line width d, the two lines can be resolved; then the resolving power is given by

$$\text{R.P.} = \frac{M}{\Delta M} = \frac{\rho}{\rho\delta + s_1 + \rho\alpha^2}. \qquad \ldots\ldots(4)$$

If the width of the exit slit at B is s_2 this is added to the denominator.

These equations determine the design of the Nier-type spectrometer tube. Thus reducing the slit widths to increase the resolution is significant only while this term dominates the denominator: the first term being small in comparison, an improvement is then possible only by eliminating the α^2 term, i.e. a second-order direction focusing.

2.3. *Second-order Directional Focusing*

Herzog has given the condition for a first-order focusing effect in the angle α when the beam does not cross the field boundaries perpendicularly which, for the case of symmetry about the OY axis with straight line field boundaries (fig. 2), can be written

$$\rho \sin\phi + 2g \cos(\phi - \epsilon)/\cos\epsilon - (g^2/\rho)\sin(\phi - 2\epsilon)/\cos^2\epsilon = 0. \qquad \ldots\ldots(5)$$

This leads directly to the geometrical construction of Stephens (1934) when $\epsilon = 0$ (i.e. for normal incidence of the beam to the field boundary). Hintenberger has extended this treatment to include a second-order focusing in α by curving the boundaries (fig. 3) to radii R_1, R_2, the condition, in addition to the above, being that

$$\rho(1/R_1 + 1/R_2) = (2\rho^2/g^2)(\rho/g + 3\tan\epsilon)\cos^3\epsilon. \qquad \ldots\ldots(6)$$

From this relation it appears that it is not possible to obtain a second-order focusing with straight line boundaries ($R_1 = R_2 = \infty$) except by having a finite value for ϵ (i.e. non-perpendicular incidence) given by

$$\rho/3g = \tan\epsilon, \qquad \ldots\ldots(7)$$

with the additional requirement that

$$\tan\epsilon = -\tfrac{1}{4}\tan\tfrac{1}{2}\phi. \qquad \ldots\ldots(8)$$

This geometry has also been derived by Kerwin from a consideration of straight-line approximations to the symmetrical boundary shape which will give ideal focusing in α, viz. the curve (fig. 3 (*d*))

$$y = x(a-x)/(\rho^2 - x^2)^{1/2}. \qquad \ldots\ldots(9)$$

The tangent drawn through the point of inflection P of this curve differs from it only in the third degree and higher, and will give a second-order focusing effect.

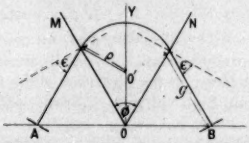


Fig. 2. Directional focusing, any angle of incidence symmetrical.

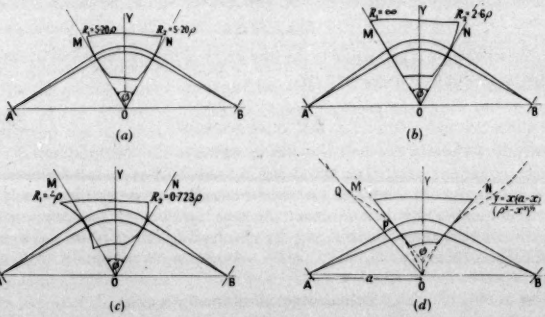


Fig. 3 (*a*)–(*d*). Second-order directional focusing.

(The conventional Nier geometry turns out to be the straight line drawn tangent to the ideal curve at Q, passing through the origin.) The dispersion of the inflection focusing arrangement is between one-third and two-thirds that of the Nier case for the same radius of curvature, so that the resolving power (neglecting the finite widths of the slits) is considerably greater and the magnet smaller in power. However, practical considerations of the contribution of the finite source width and the defocusing introduced by stray fields may render a high degree of resolution unattainable; then for little loss of resolution the aperture angle α may be increased, giving a greater beam intensity.

Another practical boundary shape, with the merit that the ion beam enters and leaves the field normally so that $\epsilon = 0$ and

$$\rho(1/R_1 + 1/R_2) = 2\rho^3/g^3 = 2\tan^3 \tfrac{1}{2}\phi, \qquad \ldots\ldots(10)$$

is that proposed by Spighel (1949). Here the boundaries are curved in the radius of the ion beam trajectory (fig. 4), giving

$$\rho/g = 1 = \tan \tfrac{1}{2}\phi; \qquad \phi = 90°. \qquad \ldots\ldots(11)$$

This is in fact a better approximation to the ideal curve than the tangent (Nier) case. Bainbridge (1947) has proposed a shape, somewhat similar, giving

$$R_1 = R_2 = \rho \tan^2 \tfrac{1}{2}\phi(\tan \tfrac{1}{2}\phi - \alpha). \qquad \ldots\ldots(12)$$

If only one pole face is available for shaping then $R_1 = \infty$ and

$$\rho/R_2 = 2\rho^3/g^3 = 2\tan^3 \tfrac{1}{2}\phi. \qquad \ldots\ldots(13)$$

For $\phi = 90°$, $R_2 = 0{\cdot}5\rho$; for $\phi = 60°$, $R_2 = 2{\cdot}6\rho$. Here the second-order aberration introduced by the straight line entrance boundary is compensated to some extent by that introduced by the curved exit boundary; and this principle of astigmatic correction is of more general application (Kerwin and Geoffrion 1949).

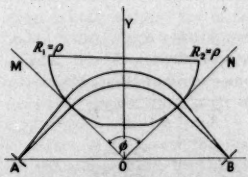


Fig. 4. Spighel's arrangement for second-order directional focusing.

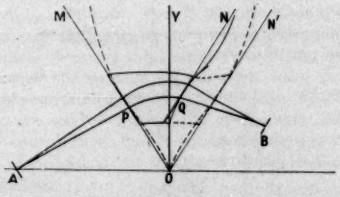


Fig. 5. Asymmetrical arrangement for high-order directional focusing.

The second-order aberration produced by a straight line approximation to the ideal boundary curve at a point where it is convex (e.g. Nier case) will be compensated by the aberration of opposite sign which is produced in so approximating the ideal curve at a point where it is concave and of equal curvature; it is found that this is just the tangent drawn to the concave part parallel to the tangent at the convex part. Since OY is the axis of symmetry, this can be arranged by displacing the right-hand quadrant of symmetry upwards by a calculable amount (fig. 5), so that the paraxial ray passing through P will pass through Q also. This results in an asymmetrical arrangement, shown in fig. 5, from which it is evident that a saving in magnet size results, in addition to the increased focusing property.

In a manner similar to that of the magnetic field case, a sector-shaped radial electrostatic field will produce a second-order aperture aberration (Cartan 1934). Hintenberger has suggested how this aberration may be matched with that produced by a homogeneous magnetic field in such a way that the second-order aperture aberration of the combination of both fields is zero. In addition, this is also possible while the field combination is such as to give a velocity focusing. This 'corrected' system is then capable of high resolution, with less restriction on the ion source.

2.4. *Stray Field Effects*

In order to assess the particular advantages of second-order focusing, and in particular to see if it results in a realizable increase in the resolving power in practice, it is necessary to examine in more detail the effect of the stray fields which necessarily occur at the magnet pole boundaries. The field does not end abruptly at the boundary edge but falls off throughout the region σ_1, σ_2 in the manner shown in fig. 6, where the ratio B_0 of the stray field to the main field is plotted along the optical axis of Herzog's idealized treatment, $S_0{}^H$. In addition, by plotting lines of equal field strength in the medial plane, the effective boundary

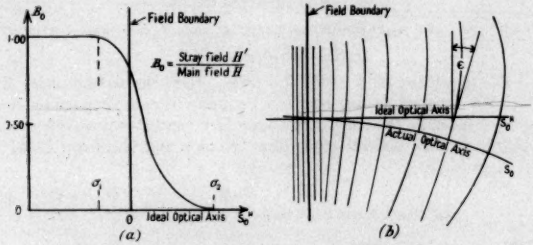


Fig. 6. Stray field effects; the boundary shape.

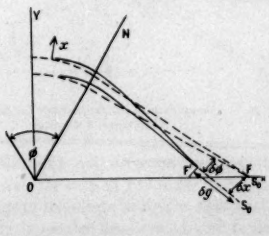


Fig. 7.

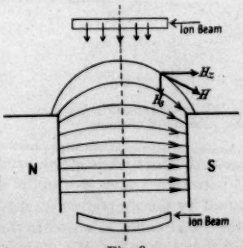


Fig. 8.

Figs. 7 and 8. Stray field effects; distortion of the axis and beam.

of the field is seen to be not the geometrical one, but one somewhat more curved (fig. 6(*b*)); the effective radius of curvature of the boundary can be computed numerically from fig. 6, using the relations

$$1/R_{\text{eff}} \simeq \int_0^1 (1/R)\, dB_0, \qquad \ldots\ldots(14)$$

where $1/R = -(\partial\epsilon/\partial x)_{x=0}$; x is measured perpendicular to the optical axis $S_0{}^H$, in the medial plane.

The gross effect of this is to bend the optical axis, first somewhat less, then more strongly, than in the idealized Herzog case: the actual optical axis S_0 is bent through an angle $\delta\phi$ and the focal point of non-axial rays is displaced a distance δg from F to F′ (fig. 7), amounts which can be estimated from the empirical curves (fig. 6) in any particular case. A good empirical correction for this can

be made in many cases by displacing the magnet bodily along the direction OY (Coggeshall 1947). However, since $\delta\phi$ may well be of the order of 1° of arc, a small correction in the value of ϕ is also to be desired in designing the pole pieces.

In addition to this shift of F, of the first order, the existence of the stray field modifies the directional focusing of those ion paths which are not paraxial, but in the medial plane. Since the effective field-boundary shape is no longer that of the pole edge, but somewhat more curved, the aperture aberration (terms involving the relevant powers of α) is multiplied by the coefficient

$$\rho(R_{\text{ideal}}/R_{\text{eff}}-1). \qquad \ldots\ldots(15)$$

Those ion paths which are not in the medial plane suffer additional defocusing; those parallel to the medial plane traverse a region between σ_1 and σ_2 when the lines of force are at an angle to the z direction (fig. 8). The smaller z component of the field, at a distance z from the medial plane, will cause less bending of the ion beam there, and will lead to a second-order aberration given by

$$d_z = -\tfrac{1}{2}z^2 \cot \tfrac{1}{2}\phi(1/R_{\text{eff}}); \qquad \ldots\ldots(16)$$

the cross section of the beam, originally rectangular, is warped (fig. 8). A consideration of (15) and (16) shows that the elimination of this is not incompatible with high-order direction focusing if R_{ideal} is infinite; such is the case when the optical axis cuts the curve of the ideal pole boundary at the point of inflection. For ion paths which cut the medial plane obliquely the Lorenz force acting on the ions, due to v_z, and the component of the field in the s-direction, viz. $\mathbf{F}(n)=\mathbf{v}_z\times\mathbf{H}_s$, will produce an additional aberration given by $d_{z(s)} \simeq \cot \tfrac{1}{2}\phi(v_z/v)z(s)$, where $z(s)=z_0+(v_z/v)s$; s is the distance measured along the optical axis, and $z=z_0$ at $s=0$. There is also a small contribution, due to the foreshortening of the velocity of the ions in the direction of the optical axis, but this is usually smaller than the above effects and can be neglected.

2.5. *Some Practical Arrangements*

The realization of a high resolving power by employing a second-order directional focusing is limited in practice by the magnitude of these defocusing effects of the second order; to minimize these R_{eff} must be large (in effect a straight line boundary is required), and the z dimension of the source slit must be restricted. One practical arrangement of obtaining this higher resolving power is that of Kerwin (1950), who has built an instrument of the inflection focusing type. A resolving power of several hundred, limited largely by slit dimensions, has been obtained with a magnet power smaller than the conventional Nier type. By narrowing down and shortening the slits, or by forming a reduced image of a source slit at the entrance slit position A (fig. 1), by means of an electrostatic lens system, the resolving power can be increased to the point where it is limited by the energy spread of the ions (a resolving power greater than 1000). A preceding velocity filter will allow still greater resolution.

An alternative arrangement proposed by Spighel (1949) employs the symmetrical case of curved field boundaries (fig. 4), to give second-order focusing. Since the ion paths are comparatively shorter, the necessity for a low pressure in the spectrometer tube is less exacting. However, the smaller value of R_{eff} may increase the stray field defocusing to such an extent that the increased directional focusing is masked. Then opening up the aperture angle α of the source will increase the sensitivity (ion beam current) with little loss in resolution.

Such a state of affairs is desirable in instruments designed to separate quantities (micrograms to milligrams) of isotopes of the elements for nuclear bombardment experiments, for spectroscopy (measuring nuclear moments), or for identifying and assigning a mass to beta activities, in conjunction with the autoradiograph technique. Bernas and Nier (1948) have suggested reshaping the pole faces of their high beam intensity instrument for this purpose. Electromagnetic separators on conventional lines have been described by Bergström *et al.* (1950) and by Koch (1948). This latter arrangement has been used to prepare targets of the noble gases by occluding the separated isotopes on the surface of Ag, Al, or Ni strips with an energy of about 60 kev; these targets remain stable for a considerable time. They can be used directly in nuclear bombarding experiments or, when necessary, the gas can be released by induction heating in a vacuum.

2.6. *Two-directional Focusing*

Ion paths parallel to the medial plane which do not cross the effective field boundary normally will suffer a force in the z direction (fig. 9(*a*)), due to the component of the field normal to the beam and the z direction. In the usual

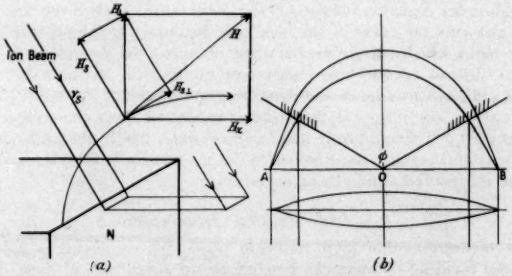


Fig. 9. Two-directional focusing.

sector field arrangement this is a defocusing effect, and can be taken account of by an increase in the z dimension of the exit slit. However, the direction of the field component in special cases may be such as to provide a focusing effect in the z direction, and this idea can be extended to provide a two-directional focusing (i.e. in both the radial and the z directions) by employing an inhomogeneous deflecting field. The form of this inhomogeneous deflecting field at a radius r, given by $H(r)=H(r_0/r)^{1/2}$, where r_0 is the medial radius of curvature, is derived from a consideration of the orbital stability in a betatron, where electrons leaving the equilibrium orbit r_0 at a small angle are refocused to it after traversing a central angle of $\pi\sqrt{2}$. A second-order treatment of the problem (Siegbahn and Svartholm 1950, Shull and Dennison 1947) shows this arrangement to be a suitable basis for a two-directional focusing device of ions from an extended source, with a large aperture angle. Such a spherical lens arrangement combines high resolution with high intensity. It would prove useful in analysing minute quantities of material, or in looking for very rare isotopes. In addition, now that improved techniques have made possible the production of high-energy

ion beams mono-energetic to a high degree (Fowler, Lauritsen and Lauritsen 1947) much useful information on detailed nuclear structure (such as narrow, closely spaced energy levels) is to be gained from the accurate analysis of the energy or momentum of the products of nuclear reactions which is made possible by an instrument of this kind. It allows the observation of a single group of particles in the presence of a background of a large flux of other particles of greater range, in particular of short range groups and residual nuclei of very short range. In cases where an incident beam of ions is the activating agent in studying prompt nuclear reactions, in order not to shift the beam on the target, it is convenient to have the target, or source, in a field-free region (fig. 9 (*b*)): this can be done by employing a sector-shaped inhomogeneous field (Judd 1950, Sugimito 1949, Snyder, Rubin and Fowler 1950), with some loss in dispersion and resolution. Because of its relatively large solid angle, this arrangement should be useful in instruments for the quantity separation of isotopes.

2.7. *The Trochotron*

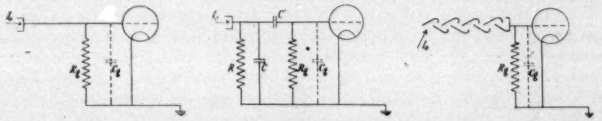
The single-focusing sector-type magnetic lens places restrictions on the type of ion source which can be employed, in that the ion beam must be approximately iso-energetic and of small angular spread. An arrangement where these restrictions disappear has been obtained by a combination of homogeneous electric and magnetic fields mutually perpendicular, in which the ions have a trochoidal path; the focal distance is independent of the initial energy or direction of the ions (Monk and Werner 1949). The advantage of a large ion beam current is offset to some extent by the fact that ions of a different m/e value can pass through a focal point on the way to their own, and by the greater difficulty in the mechanical construction and in tuning. The trochotron would appear to be most useful in fixed tuning operation, for collecting small quantities of separated isotopes, or for photographic recording, particularly so as it possesses a linear mass scale.

2.8. *Ion Beam Detection and Measurement*

Under suitable conditions the relative abundances of the isotopic components of a sample, in the region of ionization in the source, are proportional to their respective ion currents, which pass through the exit slit to the collector plate. When the resulting peak shape is flat-topped, the height of the peak gives a direct measure of the isotopic abundance; when the exit slit is too narrow to admit the whole of the beam at once, the resulting sharp peak shape must be integrated graphically to assess the total intensity. For this purpose, as also for that when two peaks are not completely resolved, a recording output meter is desirable. The problem in mass spectrometry is to measure accurately, reliably and speedily (i.e. by means of a direct-reading method) very large ratios of these ion beam currents. An upper limit of beam current ($\sim 10^{-8}$ amp) is set by beam spreading, due to the mutual electrostatic repulsion of the ions in the beam (Walcher 1943, Robinson 1949). The effect is to displace the focal point B to a new position B′, beyond B, by an amount which is calculable, depending on the ion beam current. In the usual sector field arrangement, where the collector slit is fixed in position, the space-charge effect will cause a variation in the measured isotopic ratio with a beam intensity above the figure mentioned. In any case, an intense beam requires a high gas pressure in the spectrometer, or large slit widths, or both,

resulting in a loss of resolution; or a high electron current in the ionizing region which leads to space-charge effects and considerable mass discrimination in the source itself. A lower limit is set by the noise level inherent in the ion-detecting device. Present work has been directed to increasing the signal-to-noise ratio of the detector at low signal levels.

The conventional electrometer valve method of current measurement (fig. 10(*a*)) has been improved by the development of twin electrometer valves—two electrometers in the one envelope, with a common cathode, the emission being space-charge limited—incorporated in a balanced bridge arrangement. Computations show that the potential minimum around the filament is at a distance sufficient to allow the space charge to redistribute itself almost uniformly around the cathode if a sudden change of emission occurred on one side and not on the other. Thus changes in emission current are about the same in all directions, and hence the differential change in grid current, which is what affects the zero stability, is small. Grid currents as low as 5×10^{-15} amp are obtainable, particularly with oxide-coated filaments, although this type of filament has the



(*a*) (*b*) (*c*)

Fig. 10. Ion detection arrangements.

disadvantage of long-term drifting and probably considerably greater flicker noise than a tungsten filament; d.c. amplifiers have been designed incorporating these new valves, with long-term drifts of less than 1 mV equivalent signal. The minimum measurable signal is, then, limited by the thermal noise in the grid resistor, the shot noise in the grid circuit due to the grid current I_g, and the incoming current I_0, together with a contribution from the flicker effect in the electrometer valve which it is difficult to estimate. The frequency distribution of this noise, determined by the value of R_gC_g, shows it to be mainly in the low-frequency band $0-1$ c/s (for $R_g=10^{12}$ ohms, $C_g=1$ pF), which is acceptable to the indicating meter. The signal-to-noise ratio for a bandwidth Δf is given by

$$(S/N)_E=(I_0^2R_g^2)/(4kTR_g+2eI_0R_g^2+2eI_gR_g^2)\Delta f. \quad \ldots\ldots(17)$$

With the circuit constants given above this gives a minimum detectable signal of about 5×10^{-16} amp, a figure which appears to be doubled in practice by the flicker noise.

In the case of the vibrating reed, or condenser, electrometer (Palevsky, Swank and Grenchik 1947) a d.c. signal from a high impedance source is converted to an a.c. signal across a much lower impedance grid circuit (fig. 10(*b*)) by means of a mechanically driven variable capacitance. Most of the noise in the grid circuit is now in the higher frequencies, since the product R_gC_g of the circuit is much smaller than unity; this will be undetected. However, the thermal noise, and the conduction noise due to I_0, in the high impedance source, being of low frequency, will be detected as before. Some improvement in the signal-to-noise ratio will occur at low signal levels, as if the electrometer grid current in eqn. (17) were zero. The zero stability is enhanced by a.c. rather than d.c.

amplification; synchronous detection permits a narrow band-pass about the modulating reed frequency to decrease microphonic response in the amplifier. In all an improvement by a factor 10 in S/N has been reported (Urey *et al.* 1950).

These two types of electrometer have been employed successfully in a number of arrangements either individually (e.g. Leland 1949, Inghram, Hayden and Hess 1947) or in a feedback combination to compare two simultaneously collected ion currents (Nier 1947, Urey *et al.* 1950). This latter method, although it does not allow the ion beams to fulfil the condition of traversing similar paths (Jordan and Coggeshall 1942), increases the precision of comparison of relative abundances, especially when the balance is observed over a period of time on a pen recorder, from which a good average can be estimated. By this means changes of 0·01% in the concentration of ^{13}C and ^{18}O in naturally occurring substances have been detected.

In order to measure isotopic abundance ratios of several millions, particularly in the higher mass range, where high resolution is also required (e.g. in order to set upper limits to the natural abundances of rare isotopes) a higher sensitivity

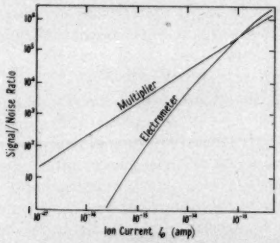


Fig. 11. Signal-to-noise ratio for signal currents.

is required; such is provided by using the secondary electron multiplier as a current pre-amplifier (fig. 10(*c*)). The incoming ion current produces a greatly increased electron current through the grid resistor R, which is much smaller than the grid resistor of the ordinary electrometer valve arrangement. The signal-to-noise ratio is given by

$$(S/N)_M=(I_0^2S^{2n}R_g^2)\Big/\left(4kTR_g+2eI_gR_g^2+2eI_0\frac{S^n(S^{n+1}-1)}{S-1}R_g^2\right)\Delta f, \quad \ldots(18)$$

where S is the gain per stage of the multiplier and n the number of stages. The first two terms of the denominator are due to the grid circuit of the electrometer, and are much smaller in general than the third, which is the noise generated by the electrometer; eqn. (18) then reduces to

$$(S/N)_M=(I_0^2R_g^2)/(2eI_0\{S/(S-1)\}R_g^2\Delta f). \quad \ldots\ldots(19)$$

The multiplier eliminates the thermal and grid current conduction noises in the grid resistor, with little increase in the shot noise of the incoming current; thus it is particularly advantageous at low signal levels. In fig. 11 eqn. (19) is plotted with $S=3$, $n=10$, $\Delta f=1$, and eqn. (17) for $R=10^{12}$ ohms; from a comparison of these it appears that below about $I_0=10^{-13}$ amp the multiplier arrangement has a higher S/N ratio.

Two arrangements of ion detection with the secondary electron multiplier have been tried. In one arrangement (Richards and Hays 1950) the ions, after passing through the exit slit, are re-accelerated by 30 kv on to a fluorescent screen, and the resulting scintillations are piped away in a quartz rod from the disturbing fields and the vacuum system to a photomultiplier. In the other method the ions are allowed to impinge upon the first dynode of the multiplier mounted directly beyond the exit slit in the spectrometer tube. Surfaces with multiplication factors greater than unity, which are relatively stable on exposure to air for some time, have been developed (Allen 1950). A Be–Cu alloy, when suitably heat-treated, will give a multiplication factor of 3 or 4 at about 400 v/stage. The noise will, in fact, be a little greater (about 20% for Be–Cu) than that given in eqn. (19), due to the fact that the distribution in the number of secondaries in the multiplication process is not Poisson above 100 v/stage. The yield of secondary electrons per ion incident on the first dynode increases linearly up to an ion energy of about 2 kev, after which it is sensibly constant (fig. 12). Thus, if the output end of the

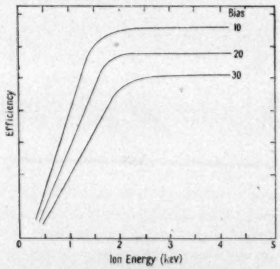


Fig. 12. Efficiency of secondary electron multiplication of Be–Cu as a function of incident ion energy.

multiplier is kept at earth potential, which is the most convenient arrangement, the ions arriving at the first dynode give a constant yield of secondaries independent of electrostatic tuning.

The multiplication factor is, however, very sensitive to surface contamination, so that it is important to keep the spectrometer tube very clean, particularly of organic material. It is also necessary to shield the multiplier very carefully from all stray fields. The multiplication factor can vary from time to time, so that the device is quite unsuitable for double collection methods; but since the input time constant is low, the spectrum can be swept back and forth a number of times rather rapidly, increasing the accuracy of measurement for single collection. In addition, the multiplication factor depends to some extent on the pressure, and the (low-frequency) noise increases rapidly as the pressure rises, presumably due to some form of plasma oscillations set up in the dynode system. The arrangement is not likely to find general use because of these limitations; but is rather of use in particular problems where the amount of material available for measurement is strictly limited. In one case it has been used to count single ions arriving at the detector, the background being as low as 10 counts/minute,

In practice the limitation of sensitivity of detection is usually set by the background impurity peaks which occur, particularly in the case of the rare earths (Leland 1950). In some cases an estimation of the impurity contribution can be made by varying the ion beam strength; however, in other cases the impurity is proportional to the sample pressure, or neighbouring weak and strong beams are insufficiently resolved, so that the multiplier cannot be used at its full sensitivity. Having regard to this, the multiplier arrangement allows the ion source to operate under optimum conditions of low pressure, small space charge, and absence of magnetic fields, thereby reducing the mass discrimination occurring in it.

For accurate absolute abundance measurements the detecting device must be linear over a considerable range. The linearity of the electron multiplier can be checked by measuring a fixed ratio, e.g. that of $^{39}K/^{40}K$, at various intensities. The method of measuring very large ratios of ion currents is, usually, to switch in, at the input of the d.c. amplifier, various grid resistors, the relative values of which can be measured and their linearity checked by timing the rate of discharge of a condenser across them in turn. Good resistors of values greater than about 10^9 ohms are difficult to obtain; apart from non-linear behaviour, they often show a polarization effect when subjected to a signal of several volts. D.C. amplifiers linear over a range of several volts are not difficult to construct.

2.9. *Ion Sources*

For absolute isotope abundance measurements the purpose of the ion source is to provide an approximately collimated, iso-energetic, and steady beam of ions, the composition of which is truly representative of the sample introduced into the ionizing region. One arrangement is to ionize a vaporous or gaseous sample by electron bombardment (Nier 1947, Thode and Graham 1947, Coggeshall 1944) in a region where the conditions of magnetic and electric field and space charge are not at all precisely defined or amenable to theoretical investigation. The existence of a magnetic field (usually of the order of 1000 gauss, coaxially with the electron beam), together with an electric field, in the source region affects to a differing degree the paths of the various ions, and tends to produce localized regions of more intense ionization. The presence of appreciable space charge, due to the ionizing electron beam, also affects the ion paths and may well be sufficient to prevent ions leaving the ionizing region unless they have a considerable initial kinetic energy. The result is to produce a discrimination between the ions of different mass, which is variable with operating conditions, including the gas pressure in the source (Careri and Nencini 1950). In addition, the solid angle accepted by the defining slits of the ion accelerating system depends upon the accelerating voltage (Coggeshall 1944). It appears desirable to reduce the space charge (i.e. the ionizing electron current), to remove the ion-repelling electrode and the magnetic field, and keep the pressure in the source region low, and to scan the spectrum by changing the magnetic deflecting field in order to eliminate as far as possible discriminating effects. This can only be done, with the present arrangement, at the expense of ion beam intensity, a condition which, with the more sensitive types of ion beam detector, has not proved restrictive, to the extent that the discrimination error is less than that introduced elsewhere, down to the low mass region. (However, ions produced (by fragmentation) with considerable initial kinetic energy will show a strong discrimination,

both in the ion source (Berry 1950), depending on the geometry and accelerating voltage, and in the analysing (spectrometer) tube (Hipple and Reese 1949) on account of the velocity component in the direction of the main field carrying them beyond the exit slit.)

In order to approximate more nearly ideal source conditions, without the disadvantage of loss of ion beam intensity, a new type of ion source has been proposed (Barnard 1949). In this the electron gun is in effect a trochotron arrangement with a weak magnetic field (about 25 gauss) in the direction of the slit lengths. The perfect focusing property of the trochotron makes it unnecessary to have beam-defining slits for the electrons, so that most of the emission from the filament is useful in producing ionization. However, the arrangement retains some of the disadvantages of the conventional source, notably the energy spreading of the ions formed in an electric field. These electron bombardment sources are also suitable for ionizing solid material evaporating from a furnace or from a filament.

Another type of ion source which has proved useful in the case of solid substances, particularly in determining the isotopic composition of the rare earths (Leland 1950, Inghram, Hayden and Hess 1947) is that of direct thermionic emission from a suitable salt (usually the oxide) 'wetted unto' a tungsten filament. The supply of ions being temperature limited, there is no space-charge discrimination. The evaporation takes place from the solid state (evaporating temperature about 1 000° C) and depends only on the ionization potential; as the source magnet can be dispensed with, the isotopic composition is truly represented in the ion beam. Preferential ion emission due to the uneven transport of ions to the emitting surface (as had been noticed, for example, in the case of K (Cook 1943)) is absent, on account of the small mass difference of the isotopes involved. The efficiency of the source (the ratio ions/atoms emitted) is constant under constant source conditions, and is high enough (the order of 10% or more) to allow measurements on very small samples.

A particular problem of ion source design has occurred in the measurement of the appearance potentials of ions and ionization probabilities, which are of some considerable importance to chemists in calculations involving bond energies and related subjects. This is usually done by electron bombardment in the conventional ion source, a calibration of the electron energy being obtained by means of test gases, usually the noble gases, admitted simultaneously with the sample under investigation. The spread of energy of the electron beam, and various other factors involving space charge and contact potentials, renders an absolute interpretation difficult. It would appear that some more clearly defined ionizing agent is required, e.g. an energy-selected electron beam of low intensity in a field-free region whose energy could be measured exactly. A means of obtaining more ideal conditions in the source region has been proposed by Fox *et al.* (1951*). Ions are formed by electron bombardment in a field-free region; the ion accelerating voltage, as a high-frequency pulse, is then applied to the ion gun plates, while the ionizing electron current is cut off. To overcome the difficulty of the energy spread of the ionizing electrons, a retarding voltage variable by an amount ΔV is applied to the electron beam in addition to its accelerating voltage; the number of ions produced by electrons in this energy range, at

* at Washington Conference on Mass Spectrometry.

any accelerating potential, is the difference in ion currents with and without the application of ΔV. Moreover, the effects of contact potential between the filament and the ionization region cancel out, when summing the retarding and accelerating potentials, at vanishing difference current to give the true appearance potential.

In those instruments designed to separate appreciable quantities of isotopes a high efficiency of ionization in the source is of more importance than the elimination of discrimination there. The chief consideration is to increase very considerably the path length of the ionizing electrons, and to this end magnetic fields are employed, either axially with respect to the accelerating tube, as in the case of Heil and Finkelstein types of source, or transversely as in the trochotron type of source proposed by Cassignol (1950), or a modification of Barnard's source. Bernas and Nier (1948) have described a new source, suitable for solid materials: a low voltage arc is struck in the vapour issuing from a furnace, and ion currents of several milliamps are obtainable, the main limitation on beam current being, as in the other types of source, one of space charge at the exit slits. All these 'discharge tube' types of source are fickle, depending critically upon the operating conditions, a small change in which may result in a considerable loss of beam intensity. The energy spread of the ion beam is also considerable, being of the order of 10 v; this, together with higher gas pressures in the spectrometer, requires high ion accelerating voltages ($\sim$50 kv). An efficiency of ionization as high as 50% (Kistemaker 1950) has been reported with proper design, and by employing gas re-circulation a very high overall efficiency is possible.

§3. NUCLEAR MEASUREMENTS

3.1. *Scope of Nuclear Measurements*

The extremely high sensitivity and precision of detection which is now possible with the mass spectrometer allows a rapid analysis of samples containing as few as 10^{13} atoms (in some cases less), with a precision such that changes in the isotopic ratios (as a result of nuclear processes) involving a few million atoms are detectable. The high flux of neutrons in the chain-reacting pile provides a source for inducing reactions on a sufficient scale in a conveniently short time. In respect of sensitivity and rapidity of measurement this method exceeds the radiochemical practice. Moreover it applies particularly to those cases in which the resultant nuclei are stable or where their radiation is sufficiently weak to escape detection by counter methods. Thus, for example, the effects of K-capture processes can be quantitatively estimated in the presence of β-emission. Even when the resultant nucleus is short-lived, providing that the extraction and analysis of the sample can take place in a time comparable with the half-life, and that this latter is known accurately, quantitative measurements on the relative yield can be undertaken.

Measurements in the neighbourhood of magic numbers are particularly interesting in view of the various theories of shell structure (Mayer 1949, Feenberg and Hammack 1949, Nordheim 1950).

3.2. *Fission Yields*

The importance of a study of the fission yields from ^{235}U lies in an elucidation of the fine structure, or perturbations, which are found to occur in the smooth mass yield curve derived on the postulate of equal charge displacement of the

light and heavy fission fragments. This divergence of the theoretical and measured fission yields has been attributed to the enhanced stability of fission fragments with a closed shell configuration (50 or 82 neutrons). This shell stabilization may cause an excited fission product, with one neutron in excess of a closed shell, to 'boil off' this loosely bound particle, either promptly or delayed, rather than decay by β-emission. In this way it depletes its mass chain yield and enhances that of one mass unit lower. The decay scheme for the primary fission products in the range mass (131–138) can be represented by table 1.

Table 1

Sn 131 → Sb 131 → Te 131 → I 131 → Xe 131

Sn 132 → Sb 132 → Te 132 → I 132 → Xe 132

n ↑

Sn 133 → Sb 133 → Te 133 → I 133 → Xe 133 → Cs 133

n ↑

Sb 134 → Te 134 → I 134 → Xe 134

n ↑

Sb 135 → Te 135 → I 135 → Xe 135 → Cs 135 → Ba 135

n ↑

Sb 136 → Te 136 → I 136 → Xe 136

n ↑

Te 137 → I 137 → Xe 137 → Cs 137 → Ba 137

n ↑

Te 138 → I 138 → Xe 138 → Cs 138 → Ba 138

β-emission → ↑ neutron emission

The isotopic analysis of the fission gases Xe and Kr from ^{235}U by means of the mass spectrometer (Thode, Macnamara and Collins 1950, Glendenin, Edwards and Coryell 1949, 1950) is particularly advantageous as the relative isotopic yields will be affected by the extent of this branching, an effect which can be measured with great precision. In many cases the isotopes are the stable end products of a disintegration chain; they are easily extracted from the uranium samples after irradiation in the pile, and rapidly transferred to the mass spectrometer for analysis. Because of this rapid handling, measurements on the yield of radioactive members can be undertaken, providing their half-lives are accurately known, and are sufficiently long to allow preparation of the sample for the spectrometer. In this way the yield of ^{133}Xe ($t_{1/2} = 5{\cdot}271 \pm 0{\cdot}002$ days) has been determined.

The measurement of the relative isotopic abundances of fission Xe has shown that, assuming a normal yield for ^{131}Xe, the yield at ^{132}Xe is normal; that at ^{133}Xe is 1·3% higher than the theoretical curve, and at ^{134}Xe 1·9% higher. At ^{136}Xe the yield is approximately normal again. A complementary yield has been noted in the case of Kr. On the basis of table 1 these data give a qualitative correlation of the primary fission yields with the theoretical curve, but more information is required on the yields of other elements (Ba, Cs, I, Ie and Sb) before definite conclusions can be drawn regarding the extent of the branching. Work is being undertaken on the yields of some isotopes of the rare earth group (Inghram, Hess and Reynolds 1949). A more rapid extraction and analysis of samples will allow shorter-lived isotopes to be included, once their half-lives have been determined accurately. Furthermore, if shell stabilization is responsible

for these fluctuations in the yield curve, then a study of the fission of ^{233}U, with fission products displaced by two mass units, should leave the fine structure unmoved (Meitner 1950).

3.3. *Half-life Determinations*

The high precision of the mass spectrometer in making relative abundance measurements provides a useful means of determining accurately and absolutely the half-life of β-activities from several hours to many years. By following in time the decay of a mass peak due to an active isotope relative to that of a stable isotope in the same sample, the value of n_t/n_0 in the decay equation $n_t = n_0 \exp(-\lambda t)$, where $\lambda t = 0{\cdot}693t/t_{1/2}$ is directly given, and an absolute value of $t_{1/2}$ can be computed with a much greater precision than in the normal counting technique values. Furthermore, the disadvantages of these counting methods, viz. the effects of contamination background, the difficulty of determining the specific activity and the errors of geometry, are eliminated. In this way Thode has found the half-life of ^{133}Xe to be $5{\cdot}271 \pm 0{\cdot}002$ days by plotting the decay in time of the mass (133) peak relative to those due to the mass (131 + 132) peaks. This, too, provides a means of making up a sample of gas of accurately known specific activity to check the efficiency of counters for absolute counting.

The mass spectrometer can be used to measure the isotopic abundance of an activity, and from the relation

$$\lambda = \frac{\text{No. of disintegrations/gramme}}{\text{No. of active atoms/gramme}}$$

the disintegration constant can be calculated after a counting experiment; this applies particularly to long-lived activities. The long-lived ^{10}Be (McMillan 1947) and ^{14}C have been measured in this way. In the latter case various values of $t_{1/2}$ have been quoted, ranging from 6 360 years to 5 513 years; the difficulty in assessing a correct value lies in the errors introduced by the various counting methods used. A precise value would be welcome as it has a direct bearing on age measurements based on ^{14}C content (Libby and Anderson 1951). A similar need is felt in the case of ^{40}K.

The question of how a double β-decay process takes place might also be established by a mass spectrometric measurement (Inghram, Washington Conf. 1951). For example, ^{130}Te decays by double β-emission to ^{130}Xe. On the basis of the Majorana theory this can be done directly without the emission of neutrinos, i.e. $^{130}Te \rightarrow {}^{130}Xe + 2\beta$. On the basis of the Dirac theory, however, this must be accompanied by neutrino emission. Thus: $^{130}Te \rightarrow {}^{130}Xe + 2\beta + 2\eta$. This leads to a much longer half-life, of about 10^{24} years, than the simpler Majorana process, with half-life about 10^{15} years. A sample of Xe gas extracted from Bi_2Te_3 mineral should show an enrichment of the ^{130}Xe isotope, depending on the value of this half-life and its geological age. A mass spectrometric analysis has been carried out on an extremely small sample, the enrichment (assuming no separation effects due to diffusion in geological time) indicating a value for the half-life of about 10^{21} years. This can be regarded as a lower limit at least, so it would appear that neutrino emission is an essential part of the decay process.

Those instruments which separate appreciable quantities of material have been used in identifying and mass-assigning new activities either by means of the auto-radiograph technique or by counting methods.

3.4. *Slow Neutron Absorption Cross Sections*

When a sample of an element is irradiated in the neutron flux of the pile its isotopic composition will undergo a change on account of the (n, γ) reactions upon the constituent nuclei. Changes of this sort in very small quantity can be measured by means of the mass spectrometer and, together with a knowledge of the flux F and time of irradiation t in the equation $n = n_0 \exp(-\sigma F t)$ governing this neutron absorption, a calculation of σ, the relative neutron cross section, for each isotope, accurate to 0·5% (or better when F and t are more accurately known) is possible.

In this way the contribution of individual isotopes to the total σ for the element can be found. For example, the high σ value in Sm is attributed to the mass (149) isotope, in Gd to the mass (155, 157) isotopes and in Nd to the mass (143) isotope; in Hg the mass (196, 199) isotopes are the principal absorbers. All these cross sections are of the order of 1000 barns.

With a more sensitive spectrometer Macnamara and Thode (1950) have undertaken similar measurements on samples of Xe and Kr of both natural and fission gas; the cross sections here are about 100 times less. Their results are shown in table 2. The low cross sections found for completed neutron shells

Table 2

Kr (Z=36)						
Mass	78	80	82	83	84	85	86
σ (barns)	—	95±15	45±15	205±10	2—0·1	15—0	2—0

Xe (Z=54)							
Mass	128	129	130	131	132	134	136
σ (barns)	5—0	45±15	5—0	120±15	5—0	5—0	5—0

at ^{86}Kr, ^{136}Xe, or shells incomplete by two neutrons at ^{84}Kr, ^{134}Xe, are to be expected on the basis of shell structure, where at shell boundaries the nucleus will show a low binding energy for an additional neutron associated with a low level density of the compound nucleus, and hence a low cross section. The low value at ^{85}Kr would seem to indicate that a closed shell less one neutron is also in a unique position. This will be noted again below.

The scope of the interpretation of these results is limited, however, by the possibility of resonance capture of thermal neutrons. It would be interesting to continue absorption cross-section measurements using unmoderated fission neutrons (~1 MEV) in order to investigate the variation of level density (which is directly connected with the cross section) with shell structure if a sufficient change in isotopic ratio to be detectable could be induced (Hughes and Sherman 1950).

3.5. *Chemical Atomic Weights; some New Rare Isotopes*

A more careful search of the vacant spaces in the periodic table with the greatly increased sensitivity now available (it being possible to detect one part in several millions in many cases) has revealed the occurrence of certain stable odd–odd nuclei, lying between two stable isobars, at neutron numbers corresponding to one less than a closed shell configuration. For example, the existence of ^{50}V (23 p, 27 n), isobaric with ^{50}Cr and ^{50}Ti, has been established by two workers (Hess and Inghram 1949, Leland 1949) to the extent of (0·23 ± 0·01)%. A check with β-counter and photographic plate reveals no detectable activity. Other

odd–odd nuclei in this region, viz. ^{46}Sc, ^{50}Mn, ^{58}Co, are well known to be active. ^{138}La (57 p, 81 n) provides another example of a stable odd–odd nucleus occurring to 0·089% (Inghram, Hayden and Hess 1947). On this basis the existence of a stable ^{208}Bi might be predicted. The suggestion is that a configuration of one less than a closed shell enjoys a peculiar stability (Broniewski 1950). In the same direction Duckworth (1949) has suggested that regularities in the location of the rarer isotopes among the heavier elements might point to the existence of other rare isotopes so far undetected. A search has revealed a ^{190}Pt isotope with an abundance of 0·012%, and upper figures have been set for the abundance of ^{118}Te (<0·0003%) and ^{129}I (<0·00003%) (Leland 1949).

The very high sensitivity of detection has also allowed an upper limit as low as 1 part in 10^6 in some cases to be put on the abundances of certain other isotopes, particularly those of the rare earths (Leland 1949, Inghram, Hayden and Hess 1947). In this work the chief limitation of sensitivity is the occurrence of background impurity peaks, the effects of which are not always to be estimated in the usual way, by varying the sample pressure. The isotopic composition of Se, Gd, Dy, Er, Eu, La, Ce and Tb has been established with high accuracy. In addition a more accurate redetermination of the isotopic composition of many of the elements has been undertaken (Nier 1950). Special care was taken to estimate any discriminating effects in the analysis; for this purpose known mixtures of the argon isotopes were made up as test samples. These data provide the material for a more accurate recalculation of the chemical atomic weights of the elements concerned, some values of which are given in table 3. It appears

Table 3

Element	Argon	Nitrogen	Oxygen	Carbon
Atomic weight calculated	39·9493 ± 3	14·00730 ± 2	16·004452 ± 7	12·0116 ± 5
International value	39·944	14·008	16·000	12·010

that there is a significant discrepancy between the chemically determined and physically determined atomic weights. However, it must be remembered that the chemical value is an average which has been arrived at by those methods in which it is intended to apply; the physically determined value is one which applies to a particular sample and, indeed, will depend on the source of this sample (see below), within appreciable limits.

3.6. *Branching Ratios*

Among the known stable nuclear species there are some 62 even–even isobaric pairs, which differ in atomic number by 2. Between these lie odd–odd nuclei of intermediate atomic number Z, which, according to present theories of β-decay, are unstable with respect to transitions either to $(Z+1)$ or to $(Z-1)$. For allowed transitions the extent of this branching will give information in the light of β-decay theory, about the position of the energy levels of the last neutron and proton in the shell structure of the Z nucleus, with respect to the levels in the resultant nuclei, having regard to selection rules governing changes in angular momentum and parity (Nordheim 1950).

Previously branching has been reported in only 12 of these 62 cases. The chief difficulty in observing it so far has been the inadequacy of experimental techniques to detect the K-capture radiation which often accompanies the transition to the $(Z-1)$ nucleus (rather than positron emission). However, if

the daughter elements $(Z+1)$ and $(Z-1)$ are produced in sufficient quantity an estimate of the relative amounts so produced will give a measure of the branching; positron emission can be detected by counting methods later. The high sensitivity of the mass spectrometer and the large yields possible in the neutron flux of the pile make this experiment a feasible one in many cases. Reynolds (1950) has in this way determined the branching ratio in ^{64}Cu, ^{80}Br, ^{82}Br and ^{128}I.

Gramme quantities of the elements in question were subjected to prolonged neutron irradiation in the pile, so that an appreciable quantity of the unstable Z nucleus was formed; this was then allowed to cool off over a period, following the reaction

$$Z(\mathrm{n},\gamma) \longrightarrow Z^* \xrightarrow{\mathrm{e}^-} Z+1$$

$$\mathrm{e}^+ \text{ or K capture} \searrow Z-1.$$

The relative amounts of these decay products were estimated, using the mass spectrometer, by means of the isotope dilution technique; this makes quantitative recovery of the daughters unnecessary. To an aliquot of the cooled-off sample weighed milligram amounts of the daughter elements, of known isotopic composition differing from that of the radiogenic daughters, were added. Then from the change in the isotopic composition of these trace daughters, on mixing with the aliquot, the relative amounts produced by either branch were determined, to give a value for the branching. Some results are given in table 4.

Table 4

Element	^{64}Cu	^{80}Br	^{82}Br	^{128}I
$\lambda\pm$	1·62	0·090	$<$0·00027	$<$0·053

$$\lambda\pm = Z \rightarrow (Z-1)/Z \rightarrow (Z+1)$$

The method is limited in application by the possibility of producing sufficient daughter elements to be measurable above the sensitivity limiting factor of impurity contribution to the mass spectrometer measurements. It would be interesting to apply it to those suitable cases resulting in a forbidden transition to the daughter nucleus: in this way the ability of the various β-decay theories to explain the process would be tried (Langer and Price 1949).

§ 4. ISOTOPIC ABUNDANCE VARIATIONS IN NATURE

4.1. *Thermodynamical Properties*

The effects of the differences in thermodynamical properties of isotopes (energy entropy and free energy) and in their properties relating to chemical kinetics (rates of reaction), to their masses (diffusion processes) and to the forces between atoms (thermal diffusion processes) leading to isotopic separations, have been studied both theoretically (Urey 1948, Bigeleisen 1949), and experimentally with the aid of the mass spectrometer. Similar processes in nature would be expected to lead to variations in the isotopic composition of the elements found there. The rarity of the heavier isotopes of the light elements precluded the quantitative study of variations in their isotopic composition until a high precision of measurement was made possible by the mass spectrometer. It has

now become apparent that small isotopic abundance variations are widespread, occurring well up in the periodic table. The interpretation of these in the light of physical and chemical processes is of considerable significance to the geologist and the biologist in explaining the history and process of formation of naturally occurring substances.

4.2. *A Paleo Temperature Scale*

The isotopic chemistry of oxygen has been studied by Urey (1948) and McCrea (1950) in some detail, in particular in regard to the equilibrium distribution of the ^{18}O isotopes in the exchange between calcium carbonate ions in solution, and water, and between calcite and water. With some simplifying assumptions McCrea produced table 5 for the fractionating factors for this isotope at various temperatures. From this it appears that there is a temperature

Table 5

Temp. (° c)	0	8	17	24
CO_3(aq) − H_2O(liq)	1·01761	1·01626	1·01486	1·01376
Calcite − H_2O(liq)	1·02548	1·02378	1·02101	1·02058

dependence of the fractionating process to the extent that a change of 1° c will change the isotopic concentration of ^{18}O by 0·02%. The measurement of these small changes is the basis of Urey's suggestion for determining paleo temperatures.

The equilibria listed in table 5 are just those in which the marine organisms lay down their shells in the ocean; moreover the ocean provides a large reservoir of oxygen which will not have changed appreciably in isotopic composition throughout geological time. A laboratory experiment on the deposition of $CaCO_3$ from a standard ocean-water sample at various temperatures provides a calibration of this exchange 'thermometer' (and a good check with theoretical deductions), in terms of the change in the isotopic composition of the oxygen, with respect to an arbitrary standard. Then a comparison with this standard, of the isotopic ratios of the oxygen from the carbonate of fossils, will give the temperature at which they were laid down. By this means a paleo temperature scale has been established over a range of 0–35°c, and seasonal variations of temperature in the growth of successive layers of shell have been noted, so that the life span and habits of the marine organism are revealed. Furthermore, samples of fossils of different ages from various parts of the world will reveal climatic conditions prevailing there at the time.

This work requires a high precision of abundance measurement, and for the purpose a special mass spectrometer was constructed along conventional lines (Urey *et al.* 1950). Simultaneous collection of the ion beams due to the two isotopes was employed, the weaker one being detected by a vibrating reed electrometer. The outputs from the two amplifiers were balanced potentiometrically one against the other, the balance being observed over a period of time on a recording meter. A special magnetically operated valve allowed rapid change from the standard to the sample under investigation. In this way changes in abundance of ^{18}O consistent to 0·01% were detectable.

The main limitation on the determination of these temperatures was an insufficient knowledge of the isotopic abundance of ^{18}O in the parent water. It appears that there is considerable variation in this through the oceans (there

being an approximate connection with the salinity). The effect of this uncertainty can be removed if measurements on the ^{18}O fractionations in phosphates, formed under the same conditions of temperature, are made; these fractionations will be smaller than in the case of carbonates.

4.3. *Isotopic Abundance Variations in Nature*

The variations in the isotopic abundances of oxygen, carbon, nitrogen and sulphur caused by exchange processes (not always fully understood) throughout geological time may prove useful in establishing the evolution of the earth's atmosphere and crust and the history of life upon it. Oxygen isotopic variations are widespread, the ^{18}O isotope being concentrated in sediments, and generally in matter which has been in solution. Rocks from deep in the earth have a low ^{18}O concentration. Variations in the $^{13}C/^{12}C$ ratio have been reported (Rankama 1948); these indicate that organic matter concentrates the lighter isotope, a fact which allows an estimation of the first appearance of life on earth. The variations of the sulphur isotopes have been measured by Thode (1950). This work indicates a spread of about 5% in the $^{32}S/^{34}S$ ratio about that of the meteoritic value; sulphites concentrate the ^{32}S, sulphates the ^{34}S. By plotting the age of the sample against the enrichment of the isotope an extrapolation will give the time at which separations first began to occur, presumably when the earth's atmosphere became oxidizing. A value of 800 million years has been quoted for this time, while a value of 500 million years has been mentioned as the probable time of the appearance of life on the planet. Work on the nitrogen isotopes (nitrogen presumably existed in the primary atmosphere of the earth as NH_3) in this direction has been lacking so far. This field of investigation in geochemistry is expanding as the precision of measurement, by means of the spectrometer, increases. Much more information is required before definite conclusions can be drawn.

4.4. *Nuclear Effects*

Much larger variations in some isotopic ratios are to be attributed to nuclear effects. The $^3He/^4He$ ratio varies by a factor of about 400, with values of 2×10^{-7} for well helium and $1{\cdot}2 \times 10^{-6}$ for atmospheric helium to as much as ten times this value for helium from some minerals, particularly those containing Li (e.g. spodumene and some of the beryls). Radioactive minerals appear to yield no ^3He (Aldrich and Nier 1948a). These ratios have been derived theoretically by Morrison and Beard (1949). They assume (*a*) that atmospheric helium occurs as a result of rock weathering to give ^4He, and cosmic-ray star processes to give ^3He by the reaction $^{14}N(n, 3\alpha)^3H$ or $^{14}N(n, ^{12}C)^3H$; (*b*) that well helium owes its ^4He to decay of Th and U, the ^3He coming from the reaction $^6Li(n, \alpha)^3H$, the neutron source being that of the (α, n) reactions of the natural α on the major rock constituents; (*c*) that this latter reaction is also responsible for the high ^3He yield in some minerals (Libby 1946). The measurement of the helium ratio is interesting in understanding nuclear geological processes. However, it presents great practical problems, in that the ratio is very small, the ^3He peak is likely to be contaminated with a background HD peak unless the resolution is high (about 500), and in many cases the amount of sample available for measurement is extremely small. In order to overcome this latter difficulty gas re-circulation can be employed, a charcoal trap cleaning the sample as it is re-cycled.

The isotopic composition of argon from potassium samples has been measured; this will show an enrichment of the ^{40}A isotope, on account of the decay of ^{40}K. With a comparison of a mineral sample and an atmospheric one, and a knowledge of the absolute amounts of radiogenic ^{40}A, and ^{40}K in the mineral, and the half-life of ^{40}K, a lower limit of the branching ratio $\lambda^{\pm}$ of 0·02 has been set (Aldrich and Nier 1948 b). A more accurate knowledge of the half-life might make this a useful method of age measurement.

REFERENCES

Aldrich, L. T., and Nier, A. O., 1948 a, *Phys. Rev.*, **74,** 1590; 1948 b, *Ibid.*, **74,** 876.

Allen, J. S., 1950, *Proc. Inst. Radio Engrs., N.Y.*, **38,** 353.

Bainbridge, K. T., 1947, *7th Solvay Congress* (Brussels).

Barber, N. F., 1933, *Proc. Leeds Phil. Soc.*, **2,** 427.

Barnard, G. P., 1949, *Nature, Lond.*, **164,** 238.

Bennett, W. H., 1950, *J. Appl. Phys.*, **21,** 143.

Bergström, I., Thulin, S., Svartholm, N., and Siegbahn, K., 1950, *Ark. Fys.*, Bd. 1, 281.

Bernas, R. H., and Nier, A. O., 1948, *Rev. Sci. Instrum.*, **19,** 895.

Berry, C. E., 1950, *Phys. Rev.*, **78,** 597.

Bigeleisen, J., 1949, *J. Chem. Phys.*, **17,** 675.

Boivin, M., 1949, *Ann. Chim. (Phys.)*, **31,** 35, etc.

Broniewski, A., 1950, *Phys. Rev.*, **77,** 846.

Careri, G., and Nencini, G., 1950, *Nuovo Cim.*, **7,** 64, *J. Chem. Phys.*, **18,** 897.

Cartan, L., 1934, *J. Phys. Radium*, **8,** 453.

Cassignol, C., 1950, *Nature, Lond.*, **166,** 234.

Coggeshall, N. D., 1944, *J. Chem. Phys.*, **12,** 19; 1947, *J. Appl. Phys.*, **18,** 855.

Cook, W. R., 1943, *Phys. Rev.*, **64,** 278.

Duckworth, H. E., 1949, *Phys. Rev.*, **75,** 1438; 1950, *Rev. Sci. Instrum.*, **21,** 54.

Eltenton, G. C., 1949, *Rev. Inst. Franç. Petrole*, **4,** 468.

Feenberg, E., and Hammack, K. C., 1949, *Phys. Rev.*, **75,** 1877.

Fowler, W. A., Lauritsen, C. C., and Lauritsen, T., 1947, *Rev. Sci. Instrum.*, **18,** 818.

Glendenin, L. E., Edwards, R. R., and Coryell, C. D., 1949, *Phys. Rev.*, **75,** 337; 1950, *Ibid.*, **77,** 755.

Goudsmit, S. A., 1948, *Phys. Rev.*, **74,** 622; 1949, *Ibid.*, **76,** 180.

Hess, D. C., and Inghram, M. G., 1949, *Phys. Rev.*, **76,** 1717.

Hintenberger, H., 1948, *Z. Naturforsch.*, **3a,** 125, 669; 1949, *Rev. Sci. Instrum.*, **20,** 748.

Hipple, J. A., and Reese, R. M., 1949, *Phys. Rev.*, **75,** 1332.

Hipple, J. A., Sommer, H., and Thomas, H. A., 1949, *Phys. Rev.*, **76,** 1877.

Hipple, J. A., and Thomas, H. A., 1949, *Phys. Rev.*, **75,** 1616.

Hughes, D. J., and Sherman, D., 1950, *Phys. Rev.*, **78,** 632.

Inghram, M. G., Hayden, R. J., and Hess, D. C., 1947, *Phys. Rev.*, **72,** 967.

Inghram, M. G., Hess, D. C., and Reynolds, J. H., 1949, *Phys. Rev.*, **76,** 1717; 1950, *Ibid.*, **79,** 271.

Jordan, E. B., and Coggeshall, N. D., 1942, *J. Appl. Phys.*, **13,** 539.

Judd, D. L., 1950, *Rev. Sci. Instrum.*, **21,** 213.

Kerwin, L., 1949, *Rev. Sci. Instrum.*, **20,** 36; 1950, *Ibid.*, **21,** 96.

Kerwin, L., and Geoffrion, C., 1949, *Rev. Sci. Instrum.*, **20,** 381.

Kistemaker, J., 1950, *Nature, Lond.*, **165,** 309.

Koch, J., 1948, *Nature, Lond.*, **161,** 566.

Lander, I., 1948, *J. Proc. Roy. Soc. N.S.W.*, **82,** 183.

Langer, L. M., and Price, H. C., 1949, *Phys. Rev.*, **76,** 641.

Leland, W. T., 1949, *Phys. Rev.*, **76,** 992, 1722; 1950, *Ibid.*, **77,** 634.

Libby, W. F., 1946, *Phys. Rev.*, **69,** 671.

Libby, W. F., and Anderson, H. L., 1951, *Phys. Rev.*, **81,** 64.

McCrea, J. M., 1950, *J. Chem. Phys.*, **18,** 849.

McMillan, E. M., 1947, *Phys. Rev.*, **72,** 591.

Macnamara, J., and Thode, H. G., 1950, *Phys. Rev.*, **80,** 296.

MATTAUCH, J., and HERZOG, R., 1934, *Z. Phys.*, **89**, 447, 786.
MAYER, M. G., 1949, *Phys. Rev.*, **75**, 1969.
MEITNER, L., 1950, *Nature, Lond.*, **165**, 561.
MONK, G. W., and WERNER, G. K., 1949, *Rev. Sci. Instrum.*, **93**, 20.
MORRISON, P., and BEARD, D. B., 1949, *Phys. Rev.*, **75**, 1332.
NEUERT, N., 1949, *Angew. Chem.*, **61**, 369.
NIER, A. O., 1947, *Rev. Sci. Instrum.*, **18**, 389; 1950, *Phys. Rev.*, **77**, 789.
NORDHEIM, L. W., 1950, *Phys. Rev.*, **78**, 294.
PALEVSKY, H., SWANK, R. K., and GRENCHIK, R., 1947, *Rev. Sci. Instrum.*, **18**, 298.
PLOCH, W., and WALCHER, W., 1950, *Z. Phys.*, **127**, 274.
RANKAMA, K., 1948, *J. Geol.*, **56**, 199.
REYNOLDS, J. H., 1950, *Phys. Rev.*, **79**, 789.
RICHARDS, P. I., and HAYS, E. E., 1950, *Rev. Sci. Instrum.*, **21**, 210.
ROBINSON, C. F., 1949, *Rev. Sci. Instrum.*, **20**, 745.
SHULL, F. B., and DENNISON, D. M., 1947, *Phys. Rev.*, **71**, 681, **72**, 256.
SIEGBAHN, K., and SVARTHOLM, N., 1950, *Proc. Phys. Soc.* A, **63**, 960.
SNYDER, C. W., RUBIN, S., FOWLER, W. A., and LAURITSEN, C. C., 1950, *Rev. Sci. Instrum.*, **21**, 852.
SPIGHEL, M., 1949, *J. Phys. Radium.*, **10**, 204.
STEPHENS, W. E., 1934, *Phys. Rev.*, **45**, 513.
SUGIMITO, A., 1949, *J. Sci. Res. Inst., Tokyo*, **43**, 223.
TAYLOR, H. S., 1948, *Amer. Phil. Soc. Proc.*, **92**, 1.
THODE, H. G., 1950, *Phys. Rev.*, **78**, 307.
THODE, H. G., and GRAHAM, R. L., 1947, *Canad. J. Res.*, **A25**, 1.
THODE, H. G., MACNAMARA, J., and COLLINS, C. B., 1950, *Phys. Rev.*, **78**, 129.
THODE, H. G., MACNAMARA, J., LOSSING, F. P., and COLLINS, C. B., 1948, *J. Amer. Chem. Soc.*, **70**, 3008.
UREY, H. C., 1948, *Science*, **108**, 489.
UREY, H. C., MCKINNEY, C. R., MCCREA, J. M., EPSTEIN, S., and ALLEN, H. A., 1950, *Rev. Sci. Instrum.*, **21**, 724.
WALCHER, W., 1943, *Z. Phys.*, **121**, 719.

RANGE–ENERGY RELATIONS

By A. E. TAYLOR
Atomic Energy Research Establishment, Harwell, Didcot, Berks.

CONTENTS

	PAGE
§ 1. Introduction	49
§ 2. Alpha-particles and protons of intermediate energies in air	50
§ 3. Slow alpha-particles and protons in air	51
§ 4. Theoretical range–energy curves for alpha-particles and protons in air	53
§ 5. Range of alpha-particles and protons in other materials and the value of the mean excitation potential	54
§ 6. Loss of energy of charged particles in gases and condensed materials	57
§ 7. Effects due to chemical combination	61
§ 8. Comparison of hydrogen and deuterium	61
(i) Molecular stopping power of hydrogen and deuterium	61
(ii) Energy loss by protons and deuterons	62
§ 9. Energy loss of very slow protons and heavy ions	62
(i) Very slow protons	62
(ii) Heavy ions	63
Acknowledgments	65
References	65

ABSTRACT. Recent work has allowed the establishment of a more accurate range–energy relation for alpha-particles and protons below 5 MEV alpha-particle energy. Measurements of the stopping power of a number of elements for protons of energies greater than 100 MEV have shown a linear relation between the mean excitation potential and the atomic number for elements of atomic number greater than that of iron. Whereas the range–energy relation for alpha-particles and protons with energies greater than 20 kev can be interpreted by quantum mechanical models of the atom, the results for very slow protons and for heavy ions and fission fragments can be interpreted by classical mechanics. Effects due to the polarization of the medium are noticed in the stopping of high-energy particles in all condensed matter and also of low-energy particles in metals.

§ 1. INTRODUCTION

THE measurement of the range of charged particles is often a very convenient way of determining their initial energy. Where the particles are gradually stopped without great deflections, a definite initial velocity will correspond to a fairly well defined range; but where there is large-angle scattering, as for electrons in heavy elements, the particles will be dispersed before losing any great part of their energy. The discussion of range–energy relations is therefore usually confined to heavy particles which have a definite range, although electrons have a fairly well defined range in the lightest substances. At first, empirical relations were used over the region covered by naturally occurring alpha-emitters. The relation between velocity V and range R in air, $R=kV^3$, was established by Geiger (1910) for ranges of alpha-particles between 3 and 7 cm. More accurate relations were sought by finding correction curves to this relation (Rutherford, Wynn-Williams, Lewis and Bowden 1933) or by altering the exponent of the velocity. For instance, the relation $R=kV^{3\cdot26}$ was found to be better for ranges greater than 5 cm (Briggs 1933). Convenient as these empirical relations were,

successful attempts have been made since then to extend the range–energy relation beyond these limits by theoretical predictions for the rate of energy loss, and these have been based on classical and quantum mechanical models of the atom. Thus, it is important to have an accurate range–energy relation not only for experimental purposes but also for the light it sheds on the appropriateness of atomistic mechanics.

With the recent advances in accelerating charged particles to high energies, it has been possible to test range–energy relations up to proton energies of 350 Mev. At still higher energies effects due to the polarization of the medium have been sought. The careful work at alpha-particle energies below 5 Mev has brought about changes in the range–energy relation in this region so that it now shows more consistent agreement with independently determined energies and ranges; at very low energies the experimental results for the energy loss have been compared with theoretical predictions. Further, the advent of fast heavily charged particles in the form of fission fragments has allowed a test of the regions of validity of classical and quantum mechanical expressions for the energy loss. It is the purpose of this review to discuss some of this recent information. Much of the discussion of the experimental results can with convenience be made in terms of the rate of energy loss.

§2. ALPHA-PARTICLES AND PROTONS OF INTERMEDIATE ENERGIES IN AIR

The range–energy relation for alpha-particles between 5 and 10 Mev was established for naturally occurring emitters and in this range little more information is now available, or indeed needed, than is given in the reviews of Livingston and Bethe (1937) and of Holloway and Livingston (1938). Recent measurements of the energy of the alpha-particles from ^{238}Pu and ^{239}Pu illustrate this. By comparing the ionization produced in argon by the Pu alpha-particle with that from Po, (energy 5·298 Mev), Jesse and Forstat (1948) arrived at $5{\cdot}140 \pm 0{\cdot}005$ Mev for the alpha-particle from ^{239}Pu and $5{\cdot}492 \pm 0{\cdot}005$ Mev for that from ^{238}Pu. Other experiments had shown that the total ionization in argon was directly proportional to the energy (Jesse, Forstat and Sadauskis 1950). By a similar method Conjeaud and Naggiar (1951) found a value $5{\cdot}134 \pm 0{\cdot}015$ Mev for ^{239}Pu. Using the ThC alpha-particle, energy 6·089 Mev, as a calibrating source in a magnetic spectrometer method, Rosenblum, Valadarès and Goldschmidt (1950) deduced a value 5·147 Mev, whereas Cranshaw and Harvey (1948) using a gridded chamber gave a value $5{\cdot}159 \pm 0{\cdot}005$ Mev. A measurement of the range in air of the alpha-particle from ^{239}Pu and ^{238}Pu (Chamberlain, Gofman, Segrè and Wahl 1947) gave 3·68 cm and 4·08 cm respectively, which correspond on the Holloway and Livingston range–energy curve (1938) to 5·15 Mev and 5·51 Mev respectively, showing the accuracy of the existing relation in this region.

Further recent evidence of the correctness of the previous range–energy curve in this region can be deduced from the measured ranges of the alpha-particles from ^{238}U and ^{234}U as $2{\cdot}702 \pm 0{\cdot}1$ and $3{\cdot}258 \pm 0{\cdot}05$ cm (Wytzes and Van der Maas 1947), which on the Holloway and Livingston curve correspond to 4·196 Mev and 4·760 Mev respectively. The energies as determined by Clark, Spencer-Palmer and Woodward (1944), using the total ionization produced in argon as a measure of the energy, were $4{\cdot}180 \pm 0{\cdot}015$ Mev. and $4{\cdot}763 \pm 0{\cdot}015$ Mev.

The range–energy relation for protons was deduced from that for the alpha-particle and was given by Livingston and Bethe (1937) as

$$R_p(E) = 1{\cdot}0072 R_\alpha(3{\cdot}971E) - 0{\cdot}20 \text{ cm} \qquad \ldots\ldots(1)$$

where $R_p(E)$ is the range of the proton of energy E, and $R_\alpha(3{\cdot}971E)$ is the range of the alpha-particle of energy $3{\cdot}971E$. The empirical constant 0·20 cm was introduced to take account of the difference in loss and capture of electrons of the ionizing particles at very low energies and was based on the work of Blackett (1932).

§3. SLOW ALPHA-PARTICLES AND PROTONS IN AIR

For energies lower than those found in natural emitting substances, reliance was placed on slowed down alpha-particles (Mano 1934 a, b) and nuclear disintegration data (Neuert 1935), and below 0·7 Mev the experimental points of Blackett and Lees (1932) were used. There was difficulty in precise energy measurement in using slowed down alpha-particles, and also little data were available from nuclear disintegrations. Holloway and Livingston (1938) measured the ionization along the path of a slow alpha-particle and deduced a range–energy relation which depends for its validity upon the assumption that the energy loss in producing an ion pair in air is independent of the energy of the alpha-particle.

Recent careful measurements of the ionization in argon and air of natural alpha groups (Jesse, Forstat and Sadauskis 1950, Jesse and Sadauskis 1949), have shown a strict proportionality between the total ionization produced and the energy of the alpha-particle in argon but not in air. Between 185 kev and 544 kev, Tunnicliffe and Ward (1952) have shown that for protons the energy w required to produce an ion pair is constant to about 1%, but that there may be a 3% difference in the ratio of the values of w for protons and alpha-particles between 2·5 Mev and 200 kev. Work by Cranshaw and Harvey (1948) and by Hanna (1950) has shown that w in argon may not be constant for low-energy alpha-particles. Jesse *et al.* measured the total ionization while the other authors have measured the electronic component of the ionization, and this difference in method may in some measure account for the discrepancy. It had been also noted by Gray (1944), in reviewing the work of Gurney (1925), that w for air varied with energy. Hence by measuring the ratio of the ionization produced in argon to that in air as a function of the residual range in air, and assuming the constancy of w in argon, it is possible to correct the energies given by Holloway and Livingston (1938). This ratio has been measured by Jesse and Sadauskis (1950) and used by them and by Bethe (1950) to give a range–energy relation for alpha-particles below 5 Mev.

This revised range–energy relation for slow alpha-particles has been compared with independently determined ranges in disintegrations where alpha-particles of known energy are emitted, and closer agreement is now obtained. Such agreement is illustrated (Jesse and Sadauskis 1950) by the range of the alpha-particle in the reaction $^6\text{Li}(n\alpha)^3\text{H}$ which has been measured as $1{\cdot}04 \pm 0{\cdot}02$ cm (Bøggild and Minnhagen 1949). This corresponds to an energy of 1·960 Mev on the Holloway and Livingston curve and an energy $2{\cdot}06 \pm 0{\cdot}03$ Mev as corrected by Jesse and Sadauskis (1950). The independently determined energy (Tollestrup, Fowler and Lauritsen 1949) is $2{\cdot}057 \pm 0{\cdot}01$ Mev. Similar agreement is shown for the reactions $^{10}\text{B}(n\alpha)$ and $^6\text{Li}(p\alpha)$. Further points have been added

by Freeman and Burcham (1951) for (pα) reactions in the region 1–2 MeV, and these are included in Table 1 which shows a comparison of the independently determined ranges and energies with the range–energy curve as revised by Jesse and Sadauskis (1950). In view of this agreement, these authors claim an accuracy of within 20 kev for energies greater than 1 MeV and 30 kev for energies less than 1 MeV.

Table 1. Comparison of Observed Ranges of Alpha-Particles from Nuclear Reactions with Ranges from the Range–Energy Relation of Jesse and Sadauskis

Reaction	Observed energy (MeV)	Ref.	Range (cm) in air at 15° c, 76 cm Hg: From energy relation	Observed	Ref.
$^{19}F(p\alpha)^{16}O$	2·113 ± 0·010	g	1·07	1·068 ± 0·008	g
$^{6}Li(n\alpha)^{3}H$	2·057 ± 0·01	e	1·037	1·04 ± 0·02	d
$^{10}B(n\alpha)^{7}Li$	1·779 ± 0·004	c	0·88	0·868 (mean)	b
$^{9}Be(p\alpha)^{6}Li$	1·774 ± 0·015	g	0·879	0·895 ± 0·008	g
$^{6}Li(p\alpha)^{3}He$	1·726	e	0·853	0·85	f
$^{19}F(p\alpha)^{16}O$	1·649 ± 0·010	g	0·811	0·812 ± 0·008	g
$^{10}B(n\alpha)^{7}Li$	1·474 ± 0·004	c	0·725	0·72 ± 0·015	a, b, h
$^{19}F(p\alpha)^{16}O$	1·326 ± 0·010	g	0·652	0·651 ± 0·008	g
$^{10}B(p\alpha)^{7}Be$	1·113 ± 0·007	g	0·551	0·559 ± 0·011	g

a, Bower, Bretscher and Gilbert (1938); b, O'Ceallaigh and Davies (1938); Bøggild (1945); c, Chao, Lauritsen and Tollestrup (1949); d, Bøggild and Minnhagen (1949); e, Tollestrup, Fowler and Lauritsen (1949); f, Neuert (1935); g, Freeman and Burcham (1951); h, Gilbert (1948).

When the alpha range–energy curve was changed it became necessary to revise the proton curve which was based on the Livingston and Bethe relation (equation (1)). Jesse and Sadauskis have given such a curve for energies between 0·3 MeV and 1·6 MeV, and have shown that the agreement with ranges and energies of protons determined from disintegration data is better than with either of the earlier curves of Livingston and Bethe (1937) or Parkinson, Herb, Bellamy and Hudson (1937). However, in such reactions as $^{14}N(np)^{14}C$ an uncertainty arises in the determination of the energy since this involves the $(n-{}^{1}H)$ mass difference. Jesse *et al.* have taken this mass difference to be 0·778 MeV while Bell and Elliott (1950) give a value 0·788 ± 0·007 MeV. These different values would correspond in the above reaction to proton energies of 0·581 MeV and 0·59 MeV respectively. The empirical constant 0·20 cm in the transformation equation (1) also offers an additional source of uncertainty.

For proton energies below 170 kev, Bethe (1950) has used the results of Crenshaw (1942), joining the two curves smoothly between 170 kev and 500 kev. Clarke and Bartholomew (1949) have also reviewed the proton range–energy curve at low energies and have given some additional points below 500 kev based on the reaction $^{2}H(\gamma n)^{1}H$ using several monochromatic γ-ray sources. The proton range was measured in a cloud chamber filled with a mixture of D_2 and D_2O vapour, and corrections were applied for the difference in stopping power in order to convert the results to mean ranges in air. The energy of the proton was calculated from the γ-ray energy using a deuteron binding energy of 2·237 ± 0·005 MeV (Bell and Elliott 1948) which has since been revised to 2·230 ± 0·007 MeV (Bell and Elliott 1950).

The possibility of obtaining further points on the curve has been suggested (Franzen, Cornog and Stephens 1948) by using those reactions where a triton is emitted, e.g. $^{6}Li(n\alpha)^{3}H$, the equivalent proton range and energy being one third of those for the triton.

The independently determined ranges and energies agree with the Bethe curve, which below 0·5 Mev shows slightly lower energies for a given range than does the curve of Jesse and Sadauskis. Table 2 shows such a comparison with the Bethe curve, and in view of this agreement Bethe claims an accuracy of within 25 kev.

Table 2. Comparison of Independently Determined Proton Ranges and Energies with Range–Energy Relation of Bethe

Reaction	Particle	Observed energy (Mev)	Ref.	Range (cm) in air at 15° c, 76 cm Hg: From energy relation	Observed	Ref.
$^{6}Li(n\alpha)^{3}H$	triton	0·912 ± 0·004	e	1·97	2·00 ± 0·02	f
$^{14}N(np)^{14}C$	proton	0·580 ± 0·008	l	1·007	0·995 ± 0·01	a, b, c
$^{3}He(np)^{3}H$	proton	0·573 ± 0·002	d	0·980	0·980 ± 0·05	b
$^{2}H(dp)^{3}H$	triton	0·296 ± 0·003	k	0·43	0·41 ± 0·03	g
$^{2}H(\gamma n)^{1}H$						
$^{24}Na\gamma$	proton	0·26 ± 0·01	k	0·378	0·44 ± 0·02	h
ThC″γ	proton	0·19 ± 0·02	k	0·27	0·23 ± 0·02	i
ThC″γ	proton	0·19 ± 0·02	k	0·27	0·155	j
ThC″γ	proton	0·19 ± 0·02	k	0·27	0·24 ± 0·01	k
$^{72}Ga\gamma$	proton	0·13 ± 0·03	k	0·19	0·18 ± 0·01	k

a, Cornog, Franzen and Stephens 1948; b, Hughes and Eggler 1948; c, Bøggild 1945; d, Taschek, Jarvis, Argo and Hemmendinger 1949; e, Tollestrup, Fowler and Lauritsen 1949; f, Bøggild and Minnhagen 1949; g, Hudspeth and Bonner 1938; h, Richardson and Emo 1938; i, Chadwick, Feather and Bretscher 1937; j, Rogers 1939; k, Clarke and Bartholomew 1949; l, Shoupp, Jennings and Sun 1949.

§4. THEORETICAL RANGE-ENERGY CURVES FOR ALPHA-PARTICLES AND PROTONS IN AIR

It is possible to separate out the terms due to nuclear interaction from the terms for electronic encounters when considering the loss of energy of charged particles in passing through matter. For the nuclear interaction the electrons are regarded as a charge distribution and the energy loss arises from collisions with the atom as a whole, while for the electronic interaction collisions with single electrons are considered. It is the electronic collisions which are mainly responsible for the energy loss of alpha-particles and protons, the contribution due to nuclear collisions being negligible. Indeed nuclear collisions only play an appreciable role in the energy loss for slowly moving fission fragments as indicated by the marked curvature of the tracks near the end of their range.

An expression for the loss of energy in electronic encounters was derived by N. Bohr (1913) using classical mechanics, whereas Bethe (1930) has given a quantum mechanical treatment which is more in accord with the results for alpha-particles and protons of intermediate and high energies. However, N. Bohr (1948) has pointed out that each treatment has its restricted region of validity, and that a classical picture is applicable for fission fragments where $2Z_1v_0/v$ is much greater than unity (v_0 is the 'velocity' of the electron in the hydrogen atom, Z_1 and v the charge and velocity of the incident particle).

The quantum mechanical relation for the energy loss was derived using Born's approximate treatment of collisions, which can be used provided the velocity of the incident particle is such that its wavelength is much less than atomic dimensions. However, allowance must be made for the electrons in those shells where the velocity of the incident particle is comparable with the atomic electron velocity. Bethe has shown that the lack of effectiveness of stopping of the K shell can be allowed for by a small correction term C_K. Further estimates for the K shell stopping power at low velocities have been made by Brown (1950), and Neufeld (1950) has evaluated the stopping power of each shell separately.

Relativistic treatments (Bethe 1932, Møller 1932) have led to additional terms in the expression for the energy loss in electronic encounters which is then given by

$$-\frac{dE}{dx}=\frac{4\pi Ne^4Z_1^2}{mv^2}Z_2\left[\ln\frac{2mv^2}{I(1-\beta^2)}-\beta^2-\frac{C_K}{Z_2}\right] \qquad \ldots\ldots(2)$$

where dE/dx is the rate of energy loss, N the number of atoms per cm^3 of the absorber, e and m the electronic charge and mass, Z_1e the charge of the incident particle, v the velocity of the incident particle, Z_2 the effective atomic number of the absorber material, I the average excitation potential of the atomic electrons of the absorber, β the ratio of v to the velocity of light, and C_K the correction term for the deficiency of stopping of the K electrons. For alpha-particles and protons in light elements, Z_2 is taken as the total number of electrons in the atom of the medium and the term C_K is then applied for the deficiency of stopping at low velocities. Values of I are derived from experiments on stopping power, although theoretical estimates have been made.

From the difference in ranges in air of the alpha-particles from ThC′ and Po, Livingston and Bethe derived a value of $I=80{\cdot}5$ ev using the above expression with $Z_2=7{\cdot}22$. Using these values, with the appropriate C_K, equation (2) fits the results of Jesse and Sadauskis down to a proton energy of 1 Mev, after which the theoretical expression gives too large a value. This is just the energy at which the velocity of the incident particle is equal to the velocity of the electron in the K shell. At lower energies a further correction term would be required for the deficiency in stopping of the L shell electrons.

§ 5. RANGE OF ALPHA-PARTICLES AND PROTONS IN OTHER MATERIALS AND THE VALUE OF THE MEAN EXCITATION POTENTIAL

While for light elements Z_2 is taken as the total number of electrons and the non-excitation of the various shell electrons is corrected for by terms such as C_K, for elements of higher atomic number the effective number of electrons as calculated by Hönl (1933) is used. For example, Livingston and Bethe have shown that the range–velocity relations for alpha-particles in gold, obtained by using for Z_2 a value 61·4 (derived from Hönl) and $I=520$ ev, agreed well with the experimental ranges and velocities determined by Rosenblum (1928). The value of I is usually chosen to give agreement with experimental values for the energy loss.

On the basis of the Fermi–Thomas model of the atom, Bloch (1933 a) showed that approximately a linear relation should hold between I and Z_2 for elements of high atomic number. The constant of proportionality a, given by Wheeler and Ladenburg (1941) on the basis of the measurements of Wilson (1941) is

11·5 ev, whereas Mano (1934 a, b) has given a value 10·5 ev for elements with atomic number greater than 20 but has shown that a is not constant for lighter elements. To obtain the value of a Mano used an expression for the energy loss given by Bloch (1933 b). Since this expression differs little from Bethe's formula when corrected for the non-excitation of the K shell electrons, it is

Table 3. Mean Excitation Potential a (ev) per Electron

Element	M, slow α-particles	B & S, 300 Mev protons	M & S, from integrated range 340 Mev protons*	B & H, 70 Mev protons*	Other references
$\frac{1}{2}H_2$	16·0				17·5(a)
in CH_2		15·6			
He	17·3				22(b)
Li	12·7	11·3			
Be		15·1	14·75		16 ± 1(c)
C solid		12·7	12·91		
$\frac{1}{2}N_2$	11·6				
Air	12·05				11·15(a)
$\frac{1}{2}O_2$	12·4				
Ne	13·2				
Al	11·9	11·5†	11·4	12·2	11·5(d)
					11·55(a)
A	10·8				
Fe		9·3			
Ni	11·6				
Cu	11·0	9·6	10·6	12·2	10·1(a)
Zn	11·3				
Kr	10·8				
Mo	10·6				
Pd	10·6				
Ag	10·4	9·1			7·77(a)
Sn	10·0	9·5			
Xe	9·8				
W		9·4			
Pt	10·1				
Au	9·9				6·6(a)
Pb	9·8	9·2	9·9	11·8	
U		9·6			

* These values of a are based upon a range corrected for multiple coulomb scattering.
† Assumed value upon which the values for other elements were based.
M, Mano 1934 a, b; B & S, Bakker and Segrè 1951; M & S, Mather and Segrè 1951; B & H, Bloembergen and van Heerden 1951; a, Livingston and Bethe 1937; b, Aron, Hoffman and Williams 1949; c, Madsen and Venkateswarlu 1948; d, Wheeler and Ladenburg 1941.

permissible to compare Mano's values of a with those of other authors. The values adopted by Livingston and Bethe (1937) for hydrogen, aluminium, air and gold differ slightly from those of Mano, but the variation of a and Z follows the same pattern, as is shown in Table 3.

In extending the Bethe formula for energy loss to higher energies, the requirement for the excitation of the inner shell electrons is reached for gold at proton energies of about 100 Mev. Thus in experiments with proton energies

in this region or greater it is possible to test the theoretical expression for energy loss. One convenient way of analysing the results is to extract the value of I, since this is the only unknown parameter. The energy loss in various materials compared with that in aluminium and copper for 345 MeV protons has been measured by Bakker and Segrè (1951), who plotted a Bragg ionization curve as a function of the thickness of absorber intercepting the proton beam. By choosing a suitable reference point for the range, different elements were substituted for the first 30 g cm^{-2} of the standard absorber and the alteration required in the remaining absorber indicated the relative stopping power. Since this 30 g cm^{-2} corresponds to an energy loss of about 75 MeV, the determinations of stopping power were for an energy of about 300 MeV. Using for aluminium the total number of electrons Z_2 and a value of $I = 150$ ev, values for the mean excitation potential per electron, a, of the other elements were obtained assuming that Z_2 for them was equal to the total number of electrons per atom. These values are included in Table 3. It will be noted that, for the elements with atomic number greater than that of iron, the constant of proportionality a is 9·4 ev. The mean range in copper (92·4 g cm^{-2}) was compared with the theoretical calculation of Aron, Hoffman and Williams (1949) who used for a a value 11·5 ev. This range would then correspond to a proton energy of 337·2 MeV, whereas using 9·6 ev for a instead of 11·5 ev would imply an initial proton energy of 341 MeV.

Mather and Segrè (1951) have used the Cerenkov radiation to determine the energy in their measurements of the range of 340 MeV protons in a number of elements from beryllium to lead. They have pointed out that due to multiple scattering in the absorber the observed range of the proton is less than the mean range. The correction for this effect for a 338·5 MeV proton in copper changes the observed range, 91·36 g cm^{-2}, to a mean range of 91·77 g cm^{-2}. Aron, Hoffman and Williams have calculated the range in various elements from the Bethe relation, equation (2), using the total number of electrons for Z_2 and $a = 11{\cdot}5$ ev. Smith (1947) has similarly published range–energy relations for air and aluminium. By a comparison with these calculated ranges, Mather and Segrè deduce the alteration in a required to reproduce their measured ranges, and these values of a are given in Table 3.

The range of protons between 35 and 120 MeV in aluminium, copper and lead has been measured by Bloembergen and van Heerden (1951) using the internal beam of the 95-inch Harvard cyclotron. The protons were deflected by coulomb scattering in a thin tungsten target and their energy determined from their $H\rho$ values to an accuracy of 0·2%. The uncertainty claimed for the mean range is less than 0·003 cm. A comparison with the theoretical calculations indicated that for aluminium the experimentally determined ranges were greater than the theoretical by $0{\cdot}9 \pm 0{\cdot}4\%$ between 56 and 76 MeV, and $1{\cdot}4 \pm 0{\cdot}1\%$ between 35 and 52 MeV. Similarly for copper the ranges were greater by $1{\cdot}0 \pm 0{\cdot}4\%$ for energies between 96 and 114 MeV, $0{\cdot}7 \pm 0{\cdot}4\%$ for energies between 73 and 90 MeV, and $1{\cdot}3 \pm 0{\cdot}4\%$ between 44 and 65 MeV. However, for lead the experimental ranges were less than the theoretical by $1{\cdot}0 \pm 0{\cdot}4\%$ in the range 50 to 115 MeV.

Making due allowance for the effect of multiple scattering on the range, Bloembergen and van Heerden derived values for the mean excitation potential using the same assumptions as Aron *et al.* The range in aluminium can be fitted by using a value $I = 159 \pm 5$ ev, while the Mather and Segrè results required a value of 150 ev, and at a lower energy (Wilson) 150 ± 5 ev was required. (Bakker

and Segrè assumed a value of 150 ev in their experiment as they were using aluminium as a standard for comparison.) However, for copper and lead there is disagreement between these several authors. Whereas Bakker and Segrè measured the differential stopping power over a limited energy range where all the shell electrons could be excited, it is doubtful whether this is so for lead in the measurements at 100 Mev (Bloembergen and van Heerden). Further, in finding a value for I by comparing the total range with that calculated from equation (2), allowance must again be made for the non-excitation of the inner shell electrons at low energies, particularly for the heavy elements. If the total number of electrons is used at all energies, then to bring about the required reduction in the energy loss a larger value of a is required, and this increase is more marked for the heavier elements. This is just the effect observed in a comparison of the results of Bakker and Segrè with those of Mather and Segrè (Table 3).

However, the value of a in the energy loss formula has less importance at high energies than at low, as is shown by the agreement of the experimentally determined ranges with the calculations of Aron *et al.* Although the values for a may differ by as much as 20%, the ranges at 100 Mev and 340 Mev agree with these calculations to within 1% or better. Thus for high energies it is possible to calculate the range from equation (2) to this accuracy by having I as the only adjustable parameter and equal to aZ_2.

§6. LOSS OF ENERGY OF CHARGED PARTICLES IN GASES AND CONDENSED MATERIALS

The expression (equation (2)) used to calculate the rate of energy loss was derived by considering the interaction of the incident particle with individual atoms. However, in a condensed medium where the electric field of the incident charged particle can extend over many atoms, the motion of the electrons in these atoms will tend to diminish that electric field. The polarization of the surrounding atoms produces a perturbation of the field of the passing particle. Such effects were suggested by Swann (1938), and Fermi (1940) has given an expression for the reduction in energy loss expected. Using only a single frequency for the binding of the electrons, which were damped in their motion, the expression for the energy loss so found depended on the dielectric constant of the material, and was such that the effect due to polarization was marked for incident particle velocities greater than the velocity of light in the material.

Although the energy loss was derived from the classical theory of collisions, for impact parameters greater than interatomic distances the polarization effect is independent of the impact parameter, and hence will have the same form in a quantum mechanical derivation. For velocities less than the velocity of light in the medium, $\mathbf{c}/\sqrt{\epsilon}$, the reduction in energy loss found by Fermi was

$$\frac{2\pi n e^4}{mv^2}\ln\epsilon \qquad \ldots\ldots(3)$$

and for velocities greater than $\mathbf{c}/\sqrt{\epsilon}$, the reduction was

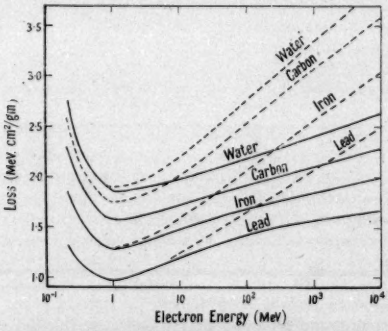
$$\frac{2\pi n e^4}{mv^2}\left[\ln\frac{\epsilon-1}{1-\beta^2}+\frac{1-\epsilon\beta^2}{\epsilon-1}\right] \qquad \ldots\ldots(4)$$

where n is the electron density, ϵ the static dielectric constant, and the other symbols have the same meaning as for equation (2). At high incident particle

energies, the energy loss given by the Bethe relation (2) increases with energy, but the polarization effect tends to reduce this and for very high energies the rate of energy loss tends to a limit determined by the electron density and frequency. Also this limit is independent of any damping of the electron vibration. The order of this reduction is illustrated by the energy loss for beta-particles in water (Fermi 1940).

Energy of beta-particles (MeV)		1	10	100
Rate of energy loss in water (MeV cm^{-1})	Fermi $\epsilon=1{\cdot}7$	1·83	1·75	1·94
	Fermi $\epsilon=1{\cdot}1$	1·92	1·91	2·09
	Bloch	1·93	2·15	2·72

Halpern and Hall (1948) have given expressions for the reduction in energy loss due to the condensation of the medium. They have used a model for the condensed medium, in which the complex dielectric constant was derived by considering more than one frequency for the bound electron, and have also taken



Theoretical ionization loss for electrons in condensed materials given by Halpern and Hall (1948). Dotted curves represent loss given by Bethe–Bloch theory, while solid curves represent loss taking polarization into account.

into account conduction electrons. The frequencies of the bound electrons for a number of substances were taken from the spectrographic and x-ray terms. Similar detailed estimates were made by Wick (1943).

Halpern and Hall have compared the results of experiments where polarization effects should be noticeable with their own calculations and with the theoretical expression of Bloch (Heitler 1944), which does not include polarization effects. The Bloch expression for fast particles, which differs slightly from the equation (2), is

$$-\frac{dE}{dx}=\frac{2\pi ne^4}{mv^2}\left[\ln\frac{mv^2T}{I^2(1-\beta^2)}+(1-\beta^2)\right] \qquad \ldots\ldots(5)$$

where n is the number of electrons per cm^3, T the maximum energy transfer in a collision, and the other symbols have the same meaning as in equation (2).

The magnitude of these effects is illustrated in the figure, taken from the paper by Halpern and Hall, and in their estimates of I in comparison with the theoretical value of $13{\cdot}5\,Z_2$ ev (Bloch 1933 a), Table 4.

It was pointed out by Fermi that where the difference between the absorption in air and that in the equivalent of a condensed substance is attributed to the half-life of the μ-meson, as in the experiments of Rossi, Hilberry and Hoag (1939), allowance must be made for polarization effects in the condensed material. Halpern and Hall have shown that in this experiment the inclusion of such effects increases the half-life by 20% of the value previously given. In comparing the absorption of mesons of energy 10^9 ev in water and lead, Pomerantz and Johnson (1940) found a relative stopping power of water to lead of $1{\cdot}72 \pm 0{\cdot}22$. The theoretical ratio based upon (5) would be 1·82 which, as Halpern and Hall have pointed out, would be reduced by 6% for polarization effects in water, the polarization effect in lead being negligible. Such a reduction would bring the theoretical relative stopping power down to 1·71, and is thus in excellent agreement with the measured value.

Similar agreement has been shown to hold for the absorption of beta-particles, but here the experiments are more difficult as corrections are needed for scattering at low energies and for radiation energy loss at high energies. However, from

Table 4. Mean Excitation Potential derived by Halpern and Hall

Material	Lead	Iron	Water	Carbon	Air	Helium
$13{\cdot}5\,Z_2$ (ev)	1100	350	135	81	95	27
I (ev)	1200	430	80	60	96	40

the measurements of Crane, Oleson and Chao (1940) on the stopping of 10 MeV beta-particles in carbon, good agreement between experiment and theory is again obtained when polarization effects are taken into account. Here corrections were applied for radiative loss. In the measurement of relative stopping power of water to carbon for 1 MeV beta-particles (Hereford 1948 a), the correction for multiple scattering was overcome by choosing the lengths of the absorbers to be the same number of radiation lengths. The measured relative stopping power of water to carbon was $1{\cdot}16 \pm 0{\cdot}025$, where, neglecting polarization effects, the ratio should be 1·08. From the curves of Halpern and Hall there will be an 8% reduction in the energy loss of 1 MeV beta-particles in carbon which will raise the theoretical value to 1·16, again giving good agreement. Similar agreement is given (Hereford 1948 b) for beta-particles up to 9 MeV.

The theory as presented by Halpern and Hall predicts that for very high energies the rate of energy loss should reach a constant value. In terms of specific ionization this value is 1·4 times minimum. Measuring in a cloud chamber the ionization of high energy beta-particles in cosmic rays, Hayward (1947) found that most of the tracks were consistent with this value, although a few high energy tracks gave much greater ionization. Hayward suggested that these might have been due to the coincidence of two or more electrons. However, in measurements of the grain density of relativistic tracks in photographic plates only about 10% rise above the minimum value is found. Messel and Ritson (1950) suggest that this may be due to the insufficient ionization energy loss in the silver bromide grain by the associated δ-rays, and to the fact that the increase for relativistic velocities is attributable to Cerenkov radiation which is undetected in these experiments. Janssens and Huybrechts (1951) have evaluated this effect.

A less direct test of the need for taking into account the polarization of the medium was considered by Whittemore and Street (1949). Here the pulse–height distribution of mesons at sea-level was measured in a silver chloride crystal, and compared with the expected distribution. Better agreement was obtained between the theory and the experimental results when polarization effects were taken into account.

From these measurements at high energy it is clear that the theory of Halpern and Hall is in large measure confirmed. As these authors point out, although they have taken a simplified model by grouping together several electron frequencies, and a more exact theory may alter the value of I, the change in dE/dx will be only a few per cent. Further confirmation is shown by the intimate connection between polarization and Cerenkov radiation, i.e. the coherent radiation emitted when the velocity of the ionizing particle is greater than the velocity of light in the medium. It is a property of the bulk of the medium rather than of its individual constituents. Indeed the relativistic increase in energy loss at very high energies is actually due to the emission of Cerenkov radiation (Schönberg 1951).

However, it is not only at high energies that effects due to the polarization of the medium will be found. As pointed out by Kramers (1947) such effects have a direct bearing on the stopping power of metals for alpha-particles of intermediate energies. Von Weizsäcker (1933) had shown that the contribution of the conduction electrons to the stopping power should be such that an increase in the electrical conductivity increases the stopping power. For instance, a change of 30% in stopping power should result from cooling lithium down to liquid helium temperatures, and for superconductors the stopping power should be infinite. In measurements of the stopping power of tin at low temperatures for polonium alpha-particles, Gerritsen (1946) found no changes greater than 0·4% for the transition into the superconductive state. Similarly, in measurements of the stopping power of aluminium from 20° K to 294° K, no change was observed within the accuracy of the experiment, whereas the theory of von Weizsäcker would give a change of 6% per conductivity electron. By considering polarization in the metal, Kramers arrives at an expression which differs little from the classical expression for energy loss. The difference corresponds to an increase of stopping power of 1 or 2%.

The effect of polarization in metals has been further studied by A. Bohr (1948), who has considered the coupling between individual encounters. Regarding the behaviour of the valency electrons as a series of oscillators corresponding to transition between the Brillouin zones, and the electrons of the inner shell as producing a negligible effect towards the polarization of the medium, estimates were made for the effective average excitation potentials I. In particular the polarization effects will be most marked for the light elements where the relative number of conduction electrons is large. A. Bohr finds for the effective average excitation potential of lithium a value of 45 ev and for beryllium a value of 60 ev, whereas the theory due to von Weizsäcker would give about 10 ev for both. Mano (1934 a) has analysed the experiments of Rosenblum (1928) for the stopping power of lithium, and shows that a value of $I = 38$ ev is required. More recent results (Howarth and King 1938) were also consistent with this value of I. Measurements of the stopping power of solid beryllium (Madsen and Venkateswarlu 1948) required a value of $I = 64 \pm 5$ ev, and additional

confirmation of this high value of I is given by the results of Bakker and Segrè (1951) and Mather and Segrè (1951), as indicated in Table 3. For separate beryllium atoms the value of I would be about 45 ev.

Even for substances other than those containing free electrons A. Bohr points out that polarization effects may not be insignificant at low energies, and Halpern and Hall have also suggested in their treatment that the frequencies of more loosely bound electrons may differ in a solid or liquid state from those in a gaseous state. This effect would be most marked for elements of small Z_2. Indeed differences have been found between the stopping power of the vapour and liquid state of a number of substances containing carbon, hydrogen and oxygen. However, the stopping power is greater for the liquid state than for the gaseous, which would be contrary to the theory of Fermi, as pointed out by Gray (1944). By comparing the ranges of RaC′ alpha-particles, Philipp (1923) found that in alcohol the integrated stopping power was 10% greater for the liquid state than for the gaseous, and that in water the stopping power was 13% greater than for the vapour. Appleyard (1949) similarly found for polonium alpha-particles a 13% greater stopping power for liquid water above the value expected. However, de Carvalho (1950) found less discrepancy in the ranges in water and ice from the expected values for RaC′ and RaF alpha-particles, and from measurements of Wilkins (1950) with emulsions containing water there also appeared to be little discrepancy.

§7. EFFECTS DUE TO CHEMICAL COMBINATION

If, as suggested in the above paragraph, a change of state may involve a modification of the electronic levels, then a similar alteration might occur with chemical combination. Any such effects should be noticeable in the comparison at low energies of the stopping powers of different gases containing the same type of atoms. However, it was noticed very early that, for gases, the stopping power was the sum of the stopping powers of their constituents (Bragg and Kleeman 1905). Recent work has confirmed this and it has been used to deduce the atomic stopping power of other elements. For instance, Hatfield and Colby (1948) have verified the additive law for a number of hydrocarbons and other gases, and have shown further that the position of the chemical bonds in the hydrocarbon does not affect the stopping power.

§8. COMPARISON OF HYDROGEN AND DEUTERIUM

(i) *Molecular Stopping Power of Hydrogen and Deuterium*

From the energy loss relations, equations (2) and (5), the atomic stopping power of hydrogen and deuterium are expected to be the same over the region of validity of those expressions. Thus only at very low energies should there be any difference which could be attributed to the differing nuclear masses. By comparing the integrated stopping power for naturally occurring alpha-particles, Schmidt, Hatfield and Lockenitz (1948) found no difference between hydrogen and deuterium. It is also evident from the results of Crenshaw (1942) that there is no difference in the stopping power down to 60 kev. It is only when the energy has fallen to values where the incident velocity is very much less than the orbital electron velocity and the loss of energy due to collisions with the atom as a whole is of significance, that the molecular stopping powers

are different. Hence down to about 1 kev proton energy, the molecular stopping powers will be the same. Differences are found, however, near the end of the ranges of fission fragments in the two gases.

(ii) *Energy Loss by Protons and Deuterons*

Using similar arguments to the above, the rates of energy loss of protons and deuterons of the same velocity are expected to be the same down to quite low energies. For speeds of 2×10^9 cm sec, i.e. proton energies of approximately 2 MeV, Laslett (1936) found that the stopping powers of the elements for deuterons did not markedly differ from Mano's values for alpha-particles, and at these velocities early experiments had indicated equal stopping powers for protons and alpha-particles. Measurements in gold and aluminium (Hall and Warshaw 1949) down to energies of 30 kev, and in air, water vapour, hydrogen, deuterium and helium down to 60 kev (Crenshaw 1942), show that the stopping powers are equal for the two particles within the experimental errors in the determinations. Indeed Hall and Warshaw estimate that the contribution due to nuclear interactions of 100 kev protons traversing a gold foil would be 0·6 kev/mg cm^{-2} in a total energy loss of 80 kev/mg cm^{-2}. This estimate is based on an equation due to N. Bohr (1941) which would give for deuterons 20% more energy loss due to nuclear collisions than for protons.

§9. ENERGY LOSS OF VERY SLOW PROTONS AND HEAVY IONS

(i) *Very Slow Protons*

When the velocity of the incident particle is much less than the orbital electron velocities, electronic encounters mainly occur without loss of energy and it is no longer possible to interpret the results by using equation (2) for energy loss. At these low velocities, that is for incident proton energies much less than 20 kev, it becomes possible to regard the electrons as a charge distribution, which may be disturbed by the encounter, but to which there is little energy loss. The main contribution then arises from nuclear collisions, and N. Bohr (1948) has shown that in such circumstances the rate of energy loss is inversely proportional to the square of the incident particle velocity. In the intermediate region the rate of energy loss becomes largely independent of the velocity.

In analysing the results for slow protons in air, French and Seidl (1951) show that the rate of energy loss in the region 27 to 57 kev (Gerthsen 1930) is proportional to the square root of the velocity, whereas in the region 4 to 50 kev (Eckardt 1930, Reusse 1932) a linear relation is more in accord with the results. For protons in the energy region 50 to 400 kev, Warshaw (1949) has measured the energy loss in a number of elements and has shown that for beryllium and aluminium his results extrapolate smoothly to the higher energy data of Madsen and Venkateswarlu (1948), and also to the theoretical curves based upon equation (2). For the other elements, copper, silver and gold, no attempt was made to draw a theoretical curve as the orbital electron velocities would be greater than the incident proton velocity. However, Neufeld (1905) has shown that by evaluating the stopping power of each electronic shell separately for proton energies greater than 300 kev by separating out the 'free' collisions from the 'resonance' collisions (N. Bohr 1948), fair agreement is obtained with the results for silver and gold. For gold the results extrapolate smoothly to the higher energy results of Huss and Madsen (1949). Further, N. Bohr (1948) states that the energy

loss in heavy elements is proportional to the cube root of the atomic number and inversely proportional to the velocity. Warshaw finds that the rate of energy loss in copper, silver and gold multiplied by $v/Z^{1/3}$ is significantly more constant than the rate of energy loss itself. For proton energies less than 20 kev Fermi and Teller (1947) have shown that the energy loss to a degenerate Fermi gas is proportional to the velocity of the incident particle. Warshaw's results extrapolated to low energies are consistent with the theoretical expression of Fermi and Teller:

$$-\frac{dE}{dx} = \frac{4}{3\pi}\frac{E}{vt_0}\ln\frac{137v_m}{c}.$$

where $t_0 = M\hbar^3/m^2e^4 = 44{\cdot}16 \times 10^{-15}\,\text{sec}^{-1}$ for protons, v_m is the Fermi velocity, M is the mass of the proton, and the other symbols have their usual meaning.

(ii) *Heavy Ions*

For heavy charged ions with velocities greater than the orbital electron velocities, the energy loss is determined by electronic encounters, but due to the high charge there will be a constant interplay of electron loss and capture. It is this which mainly governs the form of the energy loss curve and makes it differ from that for alpha-particles of comparable velocities. Thus, whereas the rate of energy loss with distance for alpha-particles increases with decreasing velocity, for heavy ions the change is much less marked, and for fission fragments it actually decreases with decreasing velocity (Lassen 1949 a). The features of the initial portion of the fission fragment track can be accounted for by considering the way in which the fragment charge varies.

In collisions it will be impossible to remove bound electrons where orbital velocities are much greater than the velocity of the charged particle, while the more loosely bound electrons are easily removed. The reverse is the case for capture of electrons, and thus the equilibrium number of bound electrons will approximately equal the number for a neutral atom for which the outermost orbital electron velocity is equal to the particle velocity. A more precise estimate for the effective charge has been given by N. Bohr (1941) as $Z_1^{1/3}v/v_o$, where Z_1 and v are the incident particle atomic number and velocity respectively and v_o is the orbital velocity in the hydrogen atom. Thus the square of the effective charge, which enters the energy loss equation (6), varies as v^2, and since the logarithmic term is approximately proportional to v, the decrease of the rate of energy loss with velocity is explained. Indeed, considering the energy loss in such electronic encounters Lassen (1949 b) has determined the effective charge as a function of velocity and found an approximately linear relationship. By using a deflection method, Lassen has also shown (1950) that the most frequent charge of the fission fragments emerging from uranium is $22e$, whereas the initial charge in argon is $15e$. The relation $Z^{\text{eff}} = Z_1^{1/3}v/v_o$ would give $20e$ and $15e$ for the uranium fission fragments with charges $Z_1 = 38$ and 54, and with velocities $6v_o$ and $4v_o$ respectively.

Lamb (1940) has identified the effective charge with the charge number of the ion formed by removal of all electrons with binding energies less than $\frac{1}{2}mv^2$. The energy loss of protons of the same velocity is then used to calculate the ranges of the two groups of uranium fission fragments giving, for an initial energy of 188 Mev and final energy of 5 Mev, 2·42 cm and 2·08 cm in agreement with experiment (Jentschke and Prankl 1939, Bøggild, Minnhagen and Nielsen 1949).

Lamb finds values of 17 e and 13 e for the initial charges. Knipp and Teller (1941) have arrived at closer estimates for the effective charge by considering the experimental results of the capture and loss of electrons by protons and alpha-particles, and also by considering the experimentally determined ranges of light element alpha recoil particles. Taking the effective charge of the ion of velocity v as determined by the relation $v_0=\gamma v$, where v_0 is the velocity of the electron which has an equal probability of being captured or lost, estimates of γ are found from the experimental results. By using a γ of 1·1 to 1·2, the theoretical range–velocity curves fit the experimental curves of the carbon, nitrogen, oxygen and fluorine alpha recoil particles. Similar estimates have been made by Brunnings and Knipp (1941) in calculating the range of neon ions in air, and here good agreement is obtained with the observed ranges up to a velocity of 4×10^8 cm/sec by using $\gamma=1{\cdot}15$, but above this velocity a smaller value of γ is required. The heavy ion becomes almost neutralized as its velocity approaches v_0, and the energy loss reaches a minimum value. However, due to the large mass of the ion, it may still possess an energy of several MeV. It is just at these velocities that nuclear collisions produce a significant contribution to the energy loss.

Since for heavy ions the ratio of the collision diameter to the wavelength is greater than unity a classical picture is applicable. For fission fragments with initial charge 20 e and energy 50 MeV, this ratio, which equals $2Z_1^{\text{eff}}e^2/\hbar v$, is approximately 10.

Using a classical picture N. Bohr (1941) has given the following expression for the velocity decrease of heavy ions,

$$\frac{1}{N}\frac{dv}{dx}=\frac{4\pi e^4}{M_1 m v^3}(Z_1^{\text{eff}})^2\Sigma_s \ln\frac{mv^3}{2\pi\nu_s e^2 Z_1^{\text{eff}}}+\frac{4\pi e^4 Z_1^{\,2}Z_2^{\,2}}{M_1M_2v^3}\ln\frac{M_1M_2v^2A_{12}^{\text{scr}}}{(M_1+M_2)Z_1Z_2e^2}. \quad (6)$$

where Z_1e, Z_2e, M_1, M_2 are the charge and mass of the fragment and gas atoms respectively, $Z_1^{\text{eff}}e$ the effective charge in electronic encounters, A_{12}^{scr} the distance between nulcei where electronic screening sets an effective limit to the action of their charges in close collisions, Σ_s is the summation over the virtual atomic oscillators, of frequency ν_s, taken with their weights, and the other symbols have their usual meaning.

The first term of this expression corresponds to electronic encounters and is valid for v greater than v_0, whereas the second term gives the loss of velocity due to nuclear collisions and is valid for v less than v_0. However, even for $v=2v_0$ the contribution due to nuclear collisions will be greater than that for electronic encounters, and thereafter the rate of velocity change will vary as v^{-3} to the end of the range.

It is due to the nuclear collisions that significant differences appear in the ranges of fission fragments in various gases. Where the electronic encounters are responsible for the energy loss and there is a linear decrease of ionization with velocity, it is possible to compare ranges by extrapolation of the linear portion of the curve to zero velocity, giving R_{ex} (Lassen 1949 a). This linear portion will be similar in various gases, but R_{ex} will be less than the ranges measured in a cloud chamber (Bøggild, Arroe and Siguorgeirsson 1946).

To compare the total ranges in various gases, it is convenient to define a reduced range as the ratio of the ranges of the fission fragment to the polonium alpha-particle range (Bøggild, Bromstrøm and Lauritsen 1941). Over the

initial portion of the range, the stopping power is nearly the same as for 4·5 MeV alpha-particles (Segrè and Wiegand 1946) so that differences in the reduced range can be attributed to nuclear collisions. Measurements in a cloud chamber (Bøggild, Arroe and Siguorgeirsson 1946) show that the reduced range increases from xenon to helium, the reduced range in helium being about 20% greater than in argon. Equation (6) predicts that the ratio of the nuclear to electronic stopping power will vary as $Z_2^{1/2}$, so that the reduced range will decrease with Z_2, in agreement with experiment. However, the reduced range in hydrogen and deuterium is less than expected upon this basis, which could be attributed to a large electronic stopping effect. This in turn means that there is an exceptionally low electron capture cross section in hydrogen and deuterium. It is also found that the range in deuterium is 6% greater than the range in hydrogen and since the electronic stopping is the same for the two this difference must be due to nuclear collisions (Bøggild, Arroe and Siguorgeirsson 1946). Assuming that the electronic stopping is inoperative for velocities less than v_0 and the nuclear stopping inoperative before this, the expected differences are 0·7 cm for the heavy group and 0·5 cm for the light group of fission fragment; such differences are found (Lassen 1949 b).

ACKNOWLEDGMENTS

I wish to thank the Director of the Atomic Energy Research Establishment for permission to publish this paper, and my many associates for their help and criticism. Thanks are also due to Professor Otto Halpern for permission to reproduce the Figure, and to several authors for permission to include their data.

REFERENCES

Appleyard, R. K., 1949, *Nature, Lond.*, **163**, 526.
Aron, W. A., Hoffman, B. G., and Williams, F. C., 1949, *U.S. Atomic Energy Commission Report*, A.E.C.U., 663.
Bakker, C. J., and Segrè, E., 1951, *Phys. Rev.*, **81**, 489.
Bell, R. E., and Elliott, L. G., 1948, *Phys. Rev.*, **74**, 1552; 1950, *Ibid.*, **79**, 282.
Bethe, H. A., 1930, *Ann. Phys., Lpz.*, **5**, 325; 1932, *Z. Phys.*, **76**, 293; 1950, *Rev. Mod. Phys.*, **22**, 213.
Blackett, P. M. S., 1932, *Proc. Roy. Soc.* A, **135**, 132.
Blackett, P. M. S., and Lees, D. S., 1932, *Proc. Roy. Soc.* A, **134**, 658.
Bloch, F., 1933 a, *Z. Phys.*, **81**, 363; 1933 b, *Ann. Phys., Lpz.*, **16**, 285.
Bloembergen, N., and van Heerden, P. J., 1951, *Phys. Rev.*, **83**, 561.
Bøggild, J. K., 1945, *K. Danske Vidensk. Selsk. Mat-fys. Medd.*, **23**, 4.
Bøggild, J. K., Arroe, O. H., and Siguorgeirsson, T., 1946, *Phys. Rev.*, **71**, 281.
Bøggild, J. K., Bromstrøm, K. J., and Lauritsen, T., 1941, *Phys. Rev.*, **59**, 275.
Bøggild, J. K., and Minnhagen, L., 1949, *Phys. Rev.*, **75**, 782.
Bøggild, J. K., Minnhagen, L., and Neilsen, O. B., 1949, *Phys. Rev.*, **76**, 988.
Bohr, A., 1948, *K. Danske Vidensk. Selsk. Mat-fys. Medd.*, **24**, 19.
Bohr, N., 1913, *Phil. Mag.*, **25**, 10; 1941, *Phys. Rev.*, **59**, 270; 1948, *K. Danske Vidensk. Selsk. Mat-fys. Medd.*, **18**, 8.
Bower, J. C., Bretscher, E., and Gilbert, C. W., 1938, *Proc. Camb. Phil. Soc.*, **34**, 290.
Bragg, W. H., and Kleeman, R., 1905, *Phil. Mag.*, **10**, 318.
Briggs, G. H., 1933, *Proc. Roy. Soc.* A, **139**, 638.
Brown, L. M., 1950, *Phys. Rev.*, **79**, 297.
Brunnings, J. H. M., and Knipp, J. K., 1941, *Phys. Rev.*, **59**, 919.
de Carvalho, H. G., 1950, *Phys. Rev.*, **78**, 330.
Chadwick, J., Feather, N., and Bretscher, E., 1937, *Proc. Roy. Soc.* A, **163**, 366.
Chamberlain, O., Gofman, J. W., Segrè, E., and Wahl, A. C., 1947, *Phys. Rev.*, **71**, 529.
Chao, C. Y., Lauritsen, C. C., and Tollestrup, A. V., 1949, *Phys. Rev.*, **76**, 586.

CLARK, F. L., SPENCER-PALMER, H. J., and WOODWARD, R. N., 1944, *British Atomic Energy Projects Report*, BR. 522.
CLARKE, R. L., and BARTHOLOMEW, G. A., 1949, *Phys. Rev.*, **76**, 146.
CONJEAUD, M., and NAGGIAR, V., 1951, *C. R. Acad. Sci., Paris*, **232**, 499.
CORNOG, I. C., FRANZEN, W., and STEPHENS, W. E., 1948, *Phys. Rev.*, **74**, 1.
CRANE, H. R., OLESON, M. L., and CHAO, K. T., 1940, *Bull. Amer. Phys. Soc.*, Feb. No. 3.
CRANSHAW, T. E., and HARVEY, J. A., 1948, *Canad. J. Res.* A, **26**, 243.
CRENSHAW, C. M., 1942, *Phys. Rev.*, **62**, 54.
ECKARDT, A., 1930, *Ann. Phys., Lpz.*, **5**, 401.
FERMI, E., 1940, *Phys. Rev.*, **57**, 485.
FERMI, E., and TELLER, E., 1947, *Phys. Rev.*, **72**, 399.
FRANZEN, W., CORNOG, I. C., and STEPHENS, W. E., 1948, *Phys. Rev.*, **74**, 1234.
FREEMAN, P. M., and BURCHAM, W. E., 1951, *Phil. Mag.*, **42**, 434.
FRENCH, A. P., and SEIDL, F. G. P., 1951, *Phil. Mag.*, **42**, 537.
GEIGER, H., 1910, *Proc. Roy. Soc.* A, **83**, 492, 505.
GERRITSEN, A. N., 1946, *Physica*, **12**, 311.
GERTHSEN, C., 1930, *Ann. Phys., Lpz.*, **5**, 657, *Phys. Z.*, **31**, 948.
GILBERT, C. W., 1948, *Proc. Camb. Phil. Soc.*, **44**, 447.
GRAY, L. H., 1944, *Proc. Camb. Phil. Soc.*, **40**, 72.
GURNEY, R. W., 1925, *Proc. Roy. Soc.* A, **107**, 332.
HALL, T. A., and WARSHAW, S. D., 1949, *Phys. Rev.*, **75**, 891.
HALPERN, O., and HALL, H., 1948, *Phys. Rev.*, **73**, 477.
HANNA, G. C., 1950, *Phys. Rev.*, **80**, 530.
HATFIELD, T. N., and COLBY, M. Y., 1948, *Phys. Rev.*, **73**, 657, *J. Franklin Inst.*, **247**, 133.
HAYWARD, E., 1947, *Phys. Rev.*, **72**, 937.
HEITLER, W., 1944, *Quantum Theory of Radiation* (Oxford : University Press).
HEREFORD, F. C., 1948 a, *Phys. Rev.*, **73**, 1123; 1948 b, *Ibid.*, **74**, 574.
HOLLOWAY, M. G., and LIVINGSTON, M. S., 1938, *Phys. Rev.*, **54**, 18.
HÖNL, H., 1933, *Z. Phys.*, **84**, 1.
HOWARTH, L. J., and KING, L. D. P., 1938, *Phys. Rev.*, **54**, 48.
HUDSPETH, E., and BONNER, T. W., 1938, *Phys. Rev.*, **54**, 308.
HUGHES, D. J., and EGGLER, C., 1948, *Phys. Rev.*, **73**, 809.
HUSS, T., and MADSEN, C. B., 1949, *Phys. Rev.*, **76**, 323.
JANSSENS, P., and HUYBRECHTS, M., 1951, *Bull. Cent. Phys. Nucl*, No. 29.
JENTSCHKE, W., and PRANKL, F., 1939, *Phys. Z.*, **40**, 706.
JESSE, W. P., and FORSTAT, H., 1948, *Phys. Rev.*, **73**, 926.
JESSE, W. P., FORSTAT, H., and SADAUSKIS, J., 1950, *Phys. Rev.*, **77**, 782.
JESSE, W. P., and SADAUSKIS, J., 1949, *Phys. Rev.*, **75**, 1110; 1950, *Ibid.*, **78**, 1.
KNIPP, J., and TELLER, E., 1941, *Phys. Rev.*, **59**, 659.
KRAMERS, H. A., 1947, *Physica*, **13**, 401.
LAMB, W. E., 1940, *Phys. Rev.*, **58**, 696.
LASLETT, L. J., 1936, *Phys. Rev.*, **50**, 391.
LASSEN, N. O., 1949 a, *Phys. Rev.*, **75**, 1762; 1949 b, *K. Danske Vidensk. Selsk. Mat-fys. Medd.*, **25**, 11; 1950, *Phys. Rev.*, **79**, 1016.
LIVINGSTON, M. S., and BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 261.
MADSEN, C. B., and VENKATESWARLU, P., 1948, *Phys. Rev.*, **74**, 648.
MANO, G., 1934 a, *Ann. Phys., Paris*, **1**, 407; 1934 b, *J. Phys. Radium*, **5**, 628.
MATHER, R., and SEGRÈ, E., 1951, *Phys. Rev.*, **84**, 191.
MESSEL, H., and RITSON, D. M., 1950, *Phil. Mag.*, **41**, 1129.
MØLLER, C., 1932, *Ann. Phys., Lpz.*, **14**, 531.
NEUERT, H., 1935, *Phys. Z.*, **36**, 629.
NEUFELD, J., 1950, *Oak Ridge National Laboratory Report.*, ORNL 884.
O'CEALLAIGH, C., and DAVIES, W. T., 1938, *Proc. Roy. Soc.* A, **167**, 81.
PARKINSON, D. B., HERB, R. G., BELLAMY, J. C., and HUDSON, C. M., 1937, *Phys. Rev.*, **52**, 75.
PHILIPP, K., 1923, *Z. Phys.*, **17**, 23.
POMERANTZ, M. A., and JOHNSON, T. H., 1940, *Bull. Amer. Phys. Soc.*, Feb., No. 4.
REUSSE, W., 1932, *Ann. Phys., Lpz.*, **15**, 256.
RICHARDSON, J. R., and EMO, L., 1938, *Phys. Rev.*, **53**, 234.

ROGERS, F. T., Jr., 1939, *Phys. Rev.*, **55**, 588.
ROSENBLUM, S., 1928, *Ann. Phys., Lpz.*, **10**, 408.
ROSENBLUM, S., VALADARÈS, M., and GOLDSCHMIDT, B., 1950, *C. R. Acad. Sci., Paris*, **230**, 638.
ROSSI, B., HILBERRY, H. V. N., and HOAG, J. B., 1939, *Phys. Rev.*, **56**, 837.
RUTHERFORD, E., WYNN-WILLIAMS, C. E., LEWIS, W. B., and BOWDEN, B. V., 1933, *Proc. Roy. Soc.* A, **139**, 617.
SCHMIDT, R. W., HATFIELD, T. N., and LOCKENITZ, A. E., 1948, *Phys. Rev.*, **73**, 652.
SCHÖNBERG, M., 1951, *Nuovo Cim.*, **8**, 159.
SHOUPP, W. E., JENNINGS, B., and SUN, K. H., 1949, *Phys. Rev.*, **75**, 1.
SEGRÈ, E., and WIEGAND, C., 1946, *Phys. Rev.*, **70**, 808.
SMITH, J. H., 1947, *Phys. Rev.*, **71**, 32.
SWANN, W. F. G., 1938, *J. Franklin Inst.*, **226**, 598.
TASCHEK, R. F., JARVIS, G. A., ARGO, H. V., and HEMMENDINGER, A., 1949, *Phys. Rev.*, **75**, 1268.
TOLLESTRUP, A. V., FOWLER, W. A., and LAURITSEN, C. C., 1949, *Phys. Rev.*, **76**, 428.
TUNNICLIFFE, P. R., and WARD, A. G., 1952, *Proc. Phys. Soc.* A, **65**, 233.
WARSHAW, R., 1949, *Phys. Rev.*, **76**, 1759.
VON WEIZSÄCKER, C. F., 1933, *Ann. Phys., Lpz.*, **5**, 17, 869.
WHEELER, J. A., and LADENBURG, R., 1941, *Phys. Rev.*, **60**, 754.
WHITTEMORE, W. L., and STREET, J. C., 1949, *Phys. Rev.*, **76**, 1786.
WICK, G. C., 1943, *Nuovo Cim.*, **1**, 302.
WILKINS, J. J., 1950, *Atomic Energy Research Establishment Report*, No. G/R 664.
WILSON, R. R., 1941, *Phys. Rev.*, **60**, 749.
WYTZES, S. A., and VAN DER MAAS, G. J., 1947, *Physica*, **13**, 49.

THE ANALYSIS OF HIGH ENERGY NEUTRON-PROTON AND PROTON-PROTON SCATTERING DATA

By R. S. CHRISTIAN*
Los Alamos Scientific Laboratory, New Mexico

CONTENTS

	PAGE
§ 1. Introduction	68
§ 2. High-energy experimental results	76
(i) Neutron sources	76
(ii) Proton sources	76
(iii) Detectors and scattering apparatus	76
§ 3. Neutron–proton scattering	77
(i) Qualitative discussion	77
(ii) Radial dependence of the potential	78
(iii) Tensor forces	82
(iv) Odd parity potentials	86
§ 4. Proton–proton scattering	88
(i) Qualitative discussion	88
(ii) Proton–proton calculations	90
§ 5. The charge-independent hypothesis	93
§ 6. Conclusions	97
References	99

ABSTRACT. The phenomenological description of the nucleon–nucleon interaction in terms of a static potential is reviewed. The scattering cross sections at various energies from 28 to 340 Mev is calculated from various assumed potential models and compared with the experimental results.

It is shown that fair agreement may be obtained for n–p scattering with a potential model having no interaction in the odd parity states and a 'long-tailed' radial dependence. The predicted total cross sections are some 20% too high, suggesting either that the potential may become repulsive at short distances or that the potential description breaks down in S states at high energies.

Proton–proton scattering results require a large fraction of the total scattering at high energies to be in the odd parity states. These odd parity forces definitely are of a non-central character. An extremely strong interaction at short distances is required which may take the form of a repulsive core in the singlet state or may be of a tensor nature in the triplet states.

A charge independent nuclear force is not indicated; however, the possibility of finding such a formulation has not been eliminated. Possible models are suggested, but at the present time more theoretical work is needed before a definite conclusion can be reached.

A summary of the knowledge of the interaction in all the various independent n–p and p–p states is reviewed.

§ 1. INTRODUCTION

A CONSIDERABLE amount of data on high-energy n–p and p–p scattering has been accumulated with the appearance of high-energy particle accelerators such as the synchro-cyclotron and the linear accelerator. n–p differential scattering cross sections have been measured over a considerable angular range with neutrons having energies ranging from 0 Mev to 260 Mev

* Now at the Institute for Advanced Study, Princeton, New Jersey.

(see fig. 1). Total cross sections over this entire energy range have also been measured (see fig. 2 and table 1). p–p differential scattering cross sections have been measured at energies ranging from 0 MeV to 340 MeV (see fig. 3).

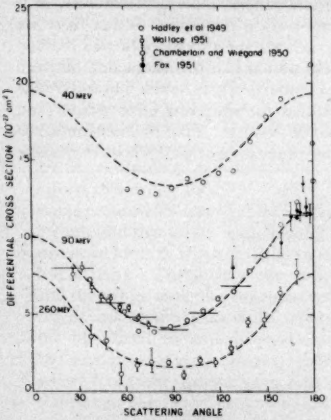


Fig. 1. Experimental n–p angular distributions. The normalization agrees with the total cross section as given in table 1. The broken line represents the best fits as determined by expansion in Legendre polynomials (see table 4). The horizontal lines at 90 MeV are the result of the cloud-chamber measurements of Brueckner *et al.* (1949). The solid points at 90 MeV near 180 degrees will be summarized by one closed triangle upon succeeding figures.

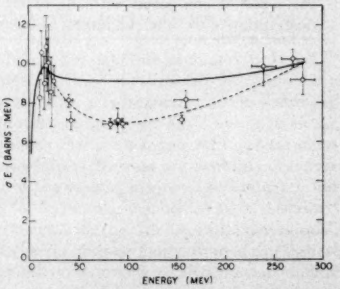


Fig. 2. (Energy) × (total cross section) plotted against energy. The experimental points below 25 MeV are from the data of Sherr (1945) and Sleator (1947) (above 40 MeV, see table 1). The full curve indicates the general characteristics of a potential model with a strong interaction at short distances such as the Yukawa potential.

Table 1

Mean energy (MeV)	Total cross section with statistical error (10^{-24} cm^2)	Detection method	Reference
41 ± 4	0·174 ± 0·010	Proton recoils	Hadley *et al.* (1949)
40 ± 4	0·203 ± 0·007	C^{13} (n, 2n) C^{11}	Hildebrand & Leith (1950)
83 ± 7	0·083 ± 0·004	C^{13} (n, 2n)C^{11}	Cook *et al.* (1947)
90 ± 3	0·079 ± 0·007	Proton recoils	Hadley *et al.* (1949)
95 ± 5	0·073 ± 0·0015	Bi fission	DeJuren & Knable (1950)
156 ± 3	0·0464 ± 0·0012	Proton recoils	Taylor *et al.* (1951)
160 ± 15	0·0512 ± 0·0026	Bi fission	DeJuren & Moyer (1951)
220 ± 15	0·041 ± 0·0024	Bi fission	DeJuren & Moyer (1951)
270 ± 15	0·038 ± 0·0015	Bi fission	DeJuren (1950)
280 ± 20	0·033 ± 0·003	Proton recoils	Fox *et al.* (1950)

n–p scattering is no longer isotropic at energies above 20 MeV, indicating the presence of scattering in other than S states. The anisotropy increases as the energy increases, indicating the appearance of scattering in states of higher and higher orbital angular momentum. p–p scattering at all energies is nearly isotropic when allowance is made for coulomb scattering at the smaller angles. The experimental cross section at 340 MeV is, however, higher than the maximum

possible on general theoretical grounds, so that scattering in other than S states is indicated here also.

A general programme would be to calculate the scattering cross section on the basis of a field theory (or perhaps even more general theory). However, as is well known, such theories are beset with theoretical difficulties so that their use for quantitative predictions is not feasible, and the most commonly used method of analysis, at present, is based upon the assumption that the interaction between two nucleons may be described by a static potential that may have a spin and charge dependence. This potential is used to calculate the scattering cross section, and the result is compared with the experimental results. This phenomenological description of the nucleon interaction has the disadvantage that it is not a relativistic

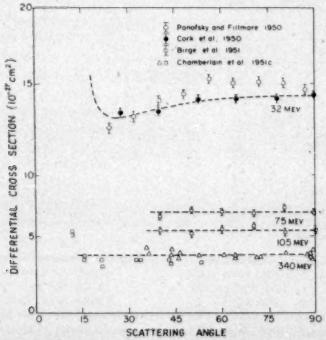


Fig. 3. Experimental p–p angular distributions. The broken curve at 32 MeV represents pure S wave nuclear scattering with a phase shift of 40° 22′. Additional experiments have been performed by Chamberlain *et al.* (1951 c) at 120 MeV in the angular range 63 to 90 degrees, at 164 MeV in the angular range 60 to 90 degrees, and at 250 MeV in the angular range 47 to 90 degrees. All these experiments show a cross section nearly independent of angle and energy with a value close to $3{\cdot}8\times10^{-27}$ cm²/sterad. Oxley *et al.* (1951) have also measured p–p scattering in the angular range 27·5 to 90 degrees at 240 MeV and found little angular variation in the cross section, the absolute value being $4{\cdot}5\pm0{\cdot}5\times10^{-27}$ cm²/sterad.

covariant procedure. Relativistic corrections may therefore arise which at least partially invalidate the direct comparison with the experimental data.

Detailed calculations, using field theories, do suggest though that the use of the relativistic momenta corresponds to calculation of the kinematical aspects of the scattering relativistically (Snyder and Marshak 1947). The dynamical effects calculated in the same way depend upon the particular theory employed and, in general, have corrections proportional to $(v/c)^2$; for example, at 90 MeV $(v/c)^2$ is 0·05, and approximately 10 per cent differences are found between cross sections calculated using the relativistic Born approximation (Møller's method) and cross sections calculated using the non-relativistic Born approximation in the case of the scalar and vector meson theories. It is possible for the corrections to be much larger, as Watson and Lepore (1949) have shown when calculations are made in higher approximations. Siegal (1951) has shown, using the method for generalizing the non-relativistic potential developed by Breit (1938), that the coefficients of the terms proportional to $(v/c)^2$ must be quite large to make corrections at 90 MeV much larger than the nominal 5 per cent figure. In fact, the coefficients

must be so large that in the low-energy region the non-relativistic theory would lose its applicability because even in the deuteron the internal motion would not be negligible.

In principle a method which is free of these objections is to determine experimentally the elements of the scattering matrix. These elements are simply related to the phase shifts in the various partial waves and are relativistic invariant quantities. There is as yet no means of predicting these phase shifts theoretically except by the use of the objectionable methods mentioned above, but their determination directly from the data would be a first step in the interpretation of the experimental results. It is, however, impossible to determine the phase shifts from the present data. This is because the neutron and proton beams (and targets) are unpolarized and consist of a mixture of singlet and triplet spin states, and vital information is lost by the averaging process. For example, at 90 MeV S, P and D states presumably account for most of the scattering (an estimation based upon the assumption that a state with angular momentum L is important in the scattering provided $kR \gtrsim L$, where k is the relative wave number and R is the range of the interaction). The singlet state, consequently, has three independent phase shifts and the triplet state nine independent phase shifts (because of possibility of tensor forces making the phase shifts different in the various magnetic substates). On the other hand, the experimental results for an unpolarized beam can be described by five independent parameters—the coefficients for the Lth Legendre polynomial, where L runs through the value zero to twice the angular momentum of the highest state taking part in the scattering. Actually in practice it is not always even possible to determine all the Legendre coefficients because the measurements extend only over a limited range of angles.

In the absence of a more comprehensive theory it is thus necessary to adopt a phenomenological approach in terms of a static potential. In the remainder of this section we shall try to find out briefly in what sense such a description is possible.

The potential function between two nucleons having only a central interaction could be assumed to have the form

$$V = (w + bP_\sigma + hP_\tau + mP_\sigma P_\tau)V(r), \qquad \ldots\ldots(1)$$

where P_σ and P_τ are the spin and charge exchange operators, respectively. The combination $-P_\sigma P_\tau$ amounts to the space exchange operator P_M for the n–p system. $V(r)$ is a radial function depending only upon the distance between the two nucleons. It, in general, falls off at large distances faster than a coulomb potential, so that the effect of the potential upon the scattering cross section comes from a finite radius.

Scattering at low energies (below 10 MeV) and the binding energy of the deuteron set limitations on the parameters w, b, h and m. Breit and his collaborators (Breit, Condon and Present 1936, Breit, Thaxton and Eisenbud 1939, Breit, Hoisington and Share 1939) have shown that p–p scattering may be described equally well by a variety of radial forms, only the range and depth of the potential being determined. More recently, Schwinger (1947 a, b) has developed an effective range theory that dispenses with the potential entirely and interprets the scattering data by means of two parameters: the Fermi scattering length a and the effective range r_0. The phase shift δ at any energy is given in the case of no coulomb potential by

$$k \cot \delta = -1/a + \tfrac{1}{2} r_0 k^2 + \ldots, \qquad \ldots\ldots(2)$$

higher order terms being negligible in the low-energy region.

These two parameters are related to the wave function for zero energy, so that the range and depth of the potential may be determined simply by solving the Schrödinger equation at zero energy of the bombarding particle. In the absence of coulomb forces, the intercept of the straight line extrapolation of the asymptotic form of the wave function with the radial axis is the scattering length and is related to the total cross section by the expression

$$\delta = 4\pi a^2/k^2. \qquad \ldots\ldots (3)$$

The effective range is related to the zero energy solution of the s-radial equation $u(r)$ by the equation:

$$r_0 = 2\int_0^\infty \left[\left(\frac{r}{a}+1\right)^2 - u^2(r)\right] dr, \qquad \ldots\ldots (4)$$

where $u(r)$ is so normalized that it asymptotically approaches $r/a+1$. Similar expressions modified by the coulomb interaction exist for the p–p system. For more details on the low-energy theory the reader is referred to a review of the p–p scattering by Jackson and Blatt (1950) and to papers by Blatt (1948), Chew and Goldberger (1949) and Bethe (1949).

Table 2

Quantity	Magnitude	Source
n–p singlet scattering length	$-23{\cdot}68 \pm 0{\cdot}06 \times 10^{-13}$ cm	Coherent scattering amplitude (Ringo & Burgy 1951); free proton cross section (Melkonian 1949)
Radius of deuteron	$4{\cdot}316 \pm 0{\cdot}004 \times 10^{-13}$ cm	γ-rays from n–p capture (Bell & Elliott 1950); photodisintegration threshold (Mobley & Laubenstein 1950); 2H–d and n–H mass difference (Taschek *et al.* 1949, Roberts and Nier 1950)
n–p triplet scattering length	$5{\cdot}388 \pm 0{\cdot}04 \times 10^{-13}$ cm	Coherent scattering amplitude (Ringo & Burgy 1951); free proton cross section (Melkonian 1949)
n–p triplet effective range	$1{\cdot}71 \pm 0{\cdot}04 \times 10^{-13}$ cm	Triplet scattering length and deuteron radius
n–p singlet effective range	$2{\cdot}7 \pm 0{\cdot}5 \times 10^{-13}$ cm	Scattering between 0 and 5 Mev (Lampi *et al.* 1950) photomagnetic disintegration, as calculated by Bethe & Longmire (1950) and Salpeter (1951)
Electric quadruple moment	$2{\cdot}73 \pm 0{\cdot}05 \times 10^{-27}$cm^2	As calculated by Nordsieck (1940)
Percentage D state admixture	3·9%	Magnetic dipole moment, neglecting relativistic effects
p–p coulomb scattering length	$-7{\cdot}67 \pm 0{\cdot}05 \times 10^{-13}$ cm	As calculated by Jackson and Blatt (1950)
p–p effective range	$2{\cdot}65 \pm 0{\cdot}07 \times 10^{-13}$ cm	As calculated by Jackson and Blatt (1950)

In table 2 are listed the effective ranges and scattering lengths, as well as other low-energy parameters which it is necessary to fit. From this table we can determine by the procedure mentioned above the parametric range necessary for the ^{3}S n–p and ^{1}S p–p states for any choice of the radial function. (It is not

necessary to determine the ^1S n–p interaction because, by our limitation to a charge independent interaction as given by eqn. (1), this interaction must be the same as for the ^1S p–p interaction. This is in good agreement with the low-energy experimental data (Schwinger 1950).) Table 3 lists the results of determining the parametric ranges for some commonly assumed radial forms. It is seen that only for the Yukawa and exponential forms are the parametric ranges even approximately equal in the singlet and triplet states, as is required by our assumption of a potential in the form of eqn. (1). With the choice of the Yukawa form, and noting that $P_\sigma(P_\tau)$ has the value 1 in a symmetric spin (charge) state, and -1 in an

Table 3

Model	p–p singlet parametric range (10^{-13} cm)	n–p triplet parametric range (10^{-13} cm)
Square	$2{\cdot}58 \pm 0{\cdot}06$	$2{\cdot}05 \pm 0{\cdot}20$
Gauss	$1{\cdot}24 \pm 0{\cdot}03$	$1{\cdot}06 \pm 0{\cdot}12$
Exponential	$0{\cdot}71 \pm 0{\cdot}02$	$0{\cdot}67 \pm 0{\cdot}08$
Yukawa	$1{\cdot}17 \pm 0{\cdot}03$	$1{\cdot}35 \pm 0{\cdot}17$

antisymmetric spin (charge) state, we have $w+b-h-m=1$ for the triplet state and $w-b-h-m=0{\cdot}69$ for the singlet state, with $V(r)=V_0\exp(-r/R)/(r/R)$, with $V_0=-67{\cdot}3$ MeV, and R (the parametric range) $=1{\cdot}18\times10^{-13}$ cm. There remain, with this form of interaction, only two constants at our disposal to fit all the high-energy data. These may conveniently be viewed as the depths of the potentials in the singlet and triplet odd parity states, the even parity state shaving been completely determined from the low-energy scattering and the assumptions implicit in eqn. (1). For orientation purposes the high-energy scattering on the basis of various commonly assumed forces is shown. The choices of the remaining constants which give these forces are:

$$\left.\begin{aligned}
&\text{Ordinary force:} && V=(0{\cdot}845+0{\cdot}155P_\sigma)V(r)\\
&\text{Exchange force:} && V=-P_\sigma P_\tau(0{\cdot}845+0{\cdot}155P_\sigma)V(r)\\
&\text{Symmetrical force:} && V=(\sigma_1\cdot\sigma_2)(\tau_1\cdot\tau_2)(0{\cdot}845+0{\cdot}155P_\sigma)V(r)\\
&\text{Even parity force:} && V=(0{\cdot}5-0{\cdot}5P_\sigma P_\tau)(0{\cdot}845+0{\cdot}155P_\sigma)V(r)
\end{aligned}\right\}\quad\ldots\ldots(5)$$

For the n–p system the ordinary force predicts scattering mainly in the forward direction, showing that, as the energy increases, small momentum transfers become increasingly more important (see fig. 4). The exchange force exhibits, on the contrary, a maximum in the backward direction, for on each collision the spin and charge of the nucleons interchange. The slight maximum in the forward direction can in this case be understood in terms of a double collision, so that the two nucleons interchange their roles twice. The symmetrical force is a mixture of part ordinary force, and, consequently, exhibits the characteristics of both the previous type forces, with the exchange type clearly predominating. The even parity force represents the limiting case of no interaction in odd parity states, and is an equal mixture of exchange and ordinary forces and hence is symmetrical about 90°. It has the further interesting consequence of predicting minimum total cross sections, since any other combination of exchange and ordinary forces gives the same cross sections in even states but has some interaction in odd states which can only lead to increased cross sections.

In the p–p system, because of the indistinguishability of the protons, the scattering is necessarily symmetrical about 90° in the centre-of-mass system. The most noticeable difference in the predicted cross sections from the various potentials now occurs in the nuclear–coulomb interference region (see fig. 5).

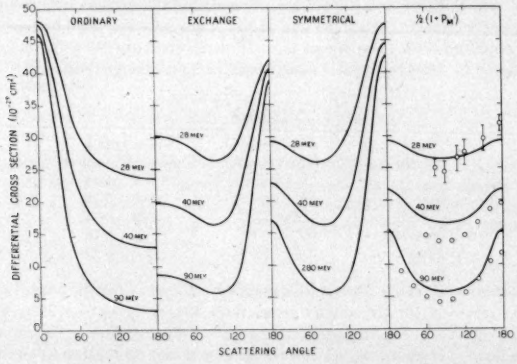


Fig. 4. n–p scattering calculated according to several commonly assumed exchange dependences.

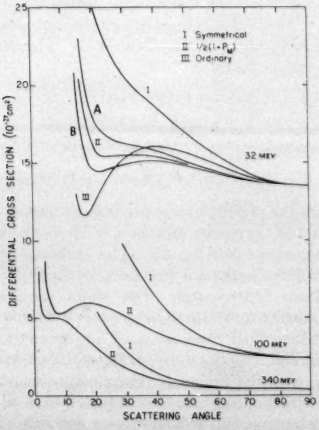


Fig. 5. p–p scattering calculated according to several commonly assumed exchange dependences. Curve III, for ordinary forces, only schematically shows the angular variation, the actual variation being much greater. The exchange dependence has not been shown either, because the large angular variation even at 32 MEV does not conveniently fit on the figure, and the variation is similar in form to that of the symmetrical force theory. Curves A and B show the effect of a $-1°$ and a $+1°$ P phase shift, respectively.

Comparing the predictions from these models with the experimental results (see figs. 1 and 3), we see that for n–p scattering reasonably good agreement is obtained with the even parity force, while the other force models disagree violently. None of the models chosen here gives any kind of agreement with the p–p data.

We have established that predictions at high energy are quite sensitive to the assumptions concerning the exchange force. It is necessary to assume a more general form for the interaction than that of eqn. (1). First, it is surely necessary to add tensor forces, since they are required to explain the quadrupole moment of the deuteron. Second, there is the logical possibility that the radial forms are different in the various states. Finally, the use of the symmetrical isotopic factor, which guarantees that the forces be charge independent, must be questioned. Lumping all these possibilities together, there are nine independent radial potentials to be determined—the singlet p–p central, triplet p–p tensor, triplet p–p central, two singlet n–p potentials of even and odd parity, two triplet n–p central potentials of even and odd parity and two triplet n–p tensor potentials of even and odd parity. (Note that the exclusion principle has operated in the p–p system to remove one half the states present in the n–p system.)

In the following sections the determination of each of these nine potentials will be considered. In particular we shall try to find some answers as to the uniqueness of the magnitude and radial form of each of the potentials. n–p scattering will be considered in § 3 and p–p scattering in § 4, the order being chosen because the predictions of the even parity force model already give fair agreement with experiment in the n–p system, whereas the theoretical predictions for the p–p cross section show the need for an essential modification of the potential interaction. Section 5 will attempt to incorporate the n–p and p–p systems into a charge independent force model. Section 2 will be reserved for a brief review of the experimental data, its main purpose being to furnish a list of references to the experimental papers.

There have been a large number of theoretical papers concerned with the interpretation of the high energy data. A summary of the theoretical predictions prior to 1948 has been given by Rosenfeld (1948, see especially sections 8.33, ad 8.33 and ad 16.22). More recent reports include papers or letters by Burhop and Yadav (1949), Yadav (1949), Castillejo and Richardson (1949), Barker and Ravenhall (1949), Hsu and Hu (1949), Rohrlich and Eisenstein (1949), Hu and Massey (1949), Demeur (1950), Christian and Hart (1950), Christian and Noyes (1950), and Jastrow (1951). The earlier papers (prior to 1950) led to uncertain (and sometimes incorrect) conclusions due primarily to the incompleteness of the experimental data at the time of publication and frequently due to the use of unjustified approximations in the calculations of the scattering. This review is accordingly based principally upon the later reports.

The methods of calculation generally used with static central potentials have been summarized by Rosenfeld (1948). The extension of the standard partial wave analysis to include the effect of non-central forces has been given by Ashkin and Wu (1948) and Case and Pais (1950). We will not consider the details in the evaluation of the phase shifts because of the diversity of methods, and the fact that different methods are more applicable in certain states than others. Among the most common methods are: step-by-step integration of the differential equation, variational procedures of the Hulthén form (for a summary see Rosenfeld 1948, chap. 5) and the Schwinger form (Lippmann and Schwinger 1950), iterational procedures based upon the integral formulation of the radial equation, as well as the

Born approximation. It is necessary to point out that the often-used Born approximation is particularly unreliable in the case of tensor forces even at the highest energies. A mathematical discussion of the validity of the Born approximation in the case of central forces has been given by Jost and Pais (1951).

§2. HIGH-ENERGY EXPERIMENTAL RESULTS

(i) *Neutron Sources*

Neutrons below 90 MeV and above 40 MeV are produced by stripping the deuteron beam of the 184 in. cyclotron at the University of California Radiation Laboratory (Helmholz *et al.* 1947). This leads to a neutron beam having a half-width of approximately 20 MeV, depending upon the exact nature of the target used. The theory of this stripping process has been given by Serber (1947), and a number of experiments have been performed showing its general verification (Chupp *et al.* 1948, Hadley *et al.* 1949).

Higher energy neutrons are produced by intercepting the circulating proton beam of the cyclotron (Kelly *et al.* 1950). Charge exchange takes place in the target leading to a wide angular distribution of neutrons in the forward direction with a mean energy of 260–270 MeV for 340 MeV protons and a half-width of 30 to 40 MeV.

Neutrons of lower energy are produced by using various nuclear reactions. For example, 28 MeV n–p scattering was measured by using the neutrons that were produced in the d(t, α)n reaction (Brolley, Coon and Fowler 1951).

(ii) *Proton Sources*

The source of nearly mono-energetic 32 MeV protons is the linear accelerator of the University of California (Bradner *et al.* 1948). The energy of the protons can be lowered to approximately 19 MeV before the loss in beam intensity makes scattering experiments impracticable (Cork 1950).

The most useful 340 MeV external proton beam was produced by scattering some of the circulating proton beam of the cyclotron into a magnetic shielding tube (Leith 1950). The beam produced in this way is extremely homogeneous in energy. Lower energy protons may be produced by putting a lithium absorber in the beam, but the intensity and collimation of the beam are reduced.

p–p scattering at 75 and 105 MeV has also been carried out on the 95 in. cyclotron at Harvard University (Birge *et al.* 1951).

(iii) *Detectors and Scattering Apparatus*

Measurements of n–p angular distributions were made by detecting the recoil protons in a telescope counter (Hadley *et al.* 1949). Low-energy protons are kept out by interposing an absorber. The energy of the incident neutron is determined from the energy of the recoil proton and the angle of scattering. The counter method is not suitable for measurement of small scattering angle since the recoil proton has too low an energy. A nuclear emulsion technique (Wallace 1951) was used in the angular range 25° to 80°, while the counter data extend from 37° to 180°. Additional measurements were made by observing the scattering in a cloud chamber (Brueckner *et al.* 1949). The agreement of all these experiments is excellent.

All the above measurements were of the relative angular distribution, so that in order to fix the absolute magnitude a knowledge of the cross section is necessary.

Several different methods have been used in order to minimize the possibility of systematic error. Perhaps the simplest of these used the counter telescope and a monitor for the beam and measured the decrease in proton recoils due to the attenuation of the neutron beam when a known absorber was placed in it. The C(n, 2n) reaction as well as the fission of bismuth have also been used as detectors. Both these detectors have the disadvantage that the variation with energy of the reaction cross section over the range of energies occurring in the neutron beam must be known. The mean effective energy of the neutron beam has been calculated by averaging an assumed $1/E$ dependence for the n–p total cross section over the experimentally determined neutron beam distribution. The C(n, 2n) reaction cross section calculated theoretically by Heckrotte and Wolff (1948) has been used in the calculation of the mean energy when carbon detectors were used. The uncertainty in determination of the mean energy makes the total cross section somewhat less reliable than the angular distribution. The experimental results and references are given in table 1.

Proton–proton scattering experiments have increased accuracy due to the homogeneity of the beam energy, the ability to measure the beam current directly with a Faraday cup, and the possibility of using coincidence counting (except at the smaller angles) to eliminate background. Detailed descriptions of the experimental arrangements are contained in the following papers: 32 MeV proportional counter method—Cork *et al.* 1950; 32 MeV photographic plate method—Panofsky and Fillmore 1950; 75 and 105 MeV scintillation counter (coincidence) method—Birge *et al.* 1951; 120 to 340 MeV scintillation counter method—Chamberlain and Wiegand 1950, Chamberlain *et al.* 1951 a, Wiegand 1950.

§ 3. NEUTRON–PROTON SCATTERING

(i) *Qualitative Discussion*

The most noteworthy feature of the n–p scattering experiments is the near symmetry about 90° evidenced particularly at 90 MeV. The expansion of the 90 MeV angular distribution in Legendre polynomials, $P_n(\theta)$,

$$4\pi\sigma(\theta) = \sigma[1 - 0{\cdot}14P_1(\theta) + 0{\cdot}68P_2(\theta) + 0{\cdot}02P_3(\theta) + 0{\cdot}11P_4(\theta)], \quad \ldots\ldots(6)$$

exhibits the smallness of the odd harmonics. (An estimation of the limits of error on the coefficients amounting to $\pm 0{\cdot}1$ was obtained by expanding several alternative curves through distributions which covered the same experimental results.)

This lack of odd harmonics in scattering indicates that the forces in the odd parity states are considerably weaker than those in the even states. A possible alternative explanation—that because the initial beam is unpolarized, one state might scatter predominantly forward while the remaining state scattered backward, the sum of the two states giving accidental symmetry—may be discounted by consideration of the total cross section. This results from the theorem that the odd state interaction can only increase the total cross section above that given by the even states alone, which by themselves predict a cross section close to that measured.

Another illustration of the smallness of the odd state interaction may be obtained by assuming that the angular distribution is the result of scattering from a single state (amounting to neglecting singlet and triplet differences) and analysing the distribution for the phase shifts. These 'average' phase shifts are: $\delta_s = 55°$, $\delta_p = -1°$, $\delta_d = 5°$, all the higher phase shifts being less than 1°. The expected

value of the P phase shift would be intermediate between that of the S and D phase shifts if the interactions were comparable, so that it can be concluded that the odd state interaction is of the order of ten times weaker than the even state interaction.

It is convenient to assume exact symmetry about 90° and study the scattering cross sections predicted by different radial forms for the potential. Each of these must, however, be adjusted to fit the low-energy data. The influence of the apparent slight asymmetry will be considered later as a small perturbation on the model best adapted to explain the major portion of the scattering.

It is also convenient to neglect tensor forces in analysing the effect of different radial forms. This can be justified since we cannot hope to find out much more concerning the radial form than some very broad classification; and the effect of a tensor potential can be approximated by an 'equivalent central potential' which has somewhat the same radial dependence as the radial dependence of the tensor potential.

(ii) *Radial Dependence of the Potential*

Using the phase shifts given above for 90 MeV, we find that the S state scattering accounts for the major fraction of the total cross section at energies up to 90 MeV. It is therefore possible to estimate the contribution from states other than S states using the experimental distribution and to determine the S state cross section with a relatively small error. A comparison of this cross section, which we shall call 'average' $\sin^2\delta_s$, with approximate statistical average of the S state cross section predicted by several different radial potentials each adjusted to fit the low-energy data, enables us to distinguish one shape characteristic. Referring to fig. 6, it is

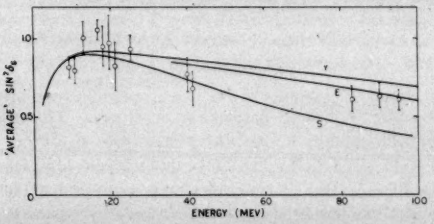


Fig. 6. S wave phase shift illustrated for a triplet effective range of $1{\cdot}65\times10^{-13}$ cm and a singlet effective range of $3{\cdot}0\times10^{-13}$ cm. Y, E, and S refer to the Yukawa, exponential and square well potentials, respectively. The experimental 'average' $\sin^2\delta_s$ is determined by the contribution of the higher partial waves as derived from the angular distribution on the subtracting basis of no spin dependence in scattering. The experimental points below 25 MeV are from the data of Sherr (1945) and Sleator (1947) (above 40 MeV, see table 1).

seen that the Yukawa potential predictions are higher, the exponential potential predictions agree closely, and the square-well potential predictions are lower than the experimental results. The interpretation of this is that since the Yukawa potential has the deepest potential near the origin, the exponential potential is less deep, and the square-well potential is the shallowest, the wavelength of the neutron is so short (1×10^{-13} cm at 90 MeV) that it can penetrate into the region of strong attraction and, consequently, the potential with the deepest hole at the origin predicts the largest S wave phase shift at high energies. This is illustrated in fig. 7, where we have plotted the various potentials and also the 90 MeV S wave function for the square and Yukawa potentials on a common radial scale.

We may expect to obtain information concerning the tail of the potential from the higher waves. At energies below 90 Mev only the D wave is appreciably perturbed (since there is little interaction in the odd parity P state). Using the spin independent model, the D wave phase shift appears to have a nearly linear energy variation, at least down to fairly low energies. In fig. 8 are illustrated the

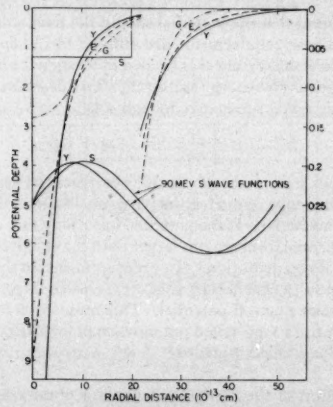


Fig. 7. A comparison with various radial dependences. The scale of the depths is arbitrary but has a common factor for all of the radial forms. Wave functions for the square and Yukawa potentials at 90 Mev are illustrated on a common but arbitrary scale, and show that the deep hole in the potential at short distances is the cause of the large S phase shift.

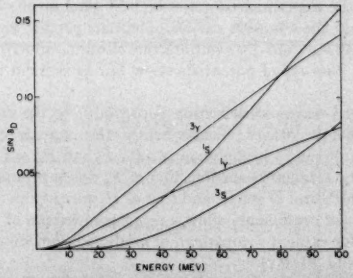


Fig. 8. D wave scattering. The quantity plotted is the sine of the D phase shift for the singlet and triplet Yukawa potential (labelled ^{1}Y and ^{3}Y) and for the singlet and triplet square well potential (labelled ^{1}S and ^{3}S).

theoretical energy variations of D phase shifts for the singlet and triplet states of a cut-off compact (square-well) potential and of a long-tailed (Yukawa) potential. It is seen that the Yukawa potential predicts a nearly linear energy variation, as is required experimentally. The interpretation is again easy to ascertain. At

40 MEV the D wave has hardly penetrated into the square well and still shows the characteristic $E^{5/2}$ energy dependence resulting because its penetration is governed primarily by the centrifugal barrier. With the Yukawa potential the D wave feels the effect of the tail at low energies, and the gradual penetration into the region of higher potential leads to more linear energy variation in the phase shift from 10 MEV on.

Thus we have found that the potential should not have a deep hole and should have a long tail. These requirements are difficult to obtain with one potential since the deep hole at small distances is needed to balance the long tail and give the proper effective range for low-energy scattering. To show this, recall the definition of the effective range given previously by eqn. (4),

$$r_0 = 2\int_0^\infty \left[\left(\frac{r}{a}+1\right)^2 - u(r)^2\right] dr.$$

The integrand differs from zero only inside the potential region and is positive everywhere, since the range of nuclear forces is smaller than the scattering length and the potential is attractive. If the potential has a long tail, contributions to the value of the integral come from regions larger than the effective range, while if it is cut off there will be no contribution. In order to maintain the effective range for the long-tailed potential it is necessary that $u(r)$ remains larger inside the effective range than in the case of a cut-off potential. This means that for small r, $u(r)$ must approach zero faster for a long-tailed potential than for the cut-off potential and, consequently, the long-tailed potential shows a stronger interaction at short distances.

A rapid comparison of the various potentials is obtained from the predicted values for the ratio $\sigma(180°)/\sigma(90°)$ (which in the absence of the odd parity states is the same as the ratio $\sigma(0°)/\sigma(90°)$). For the square potential we find the value of this ratio is 1·1 at 40 MEV and 4·2 at 90 MEV. The exponential and Yukawa potentials predict 1·3 and 1·4 at 40 MEV, and 4·0 and 3·2 at 90 MEV, respectively. Comparing with the experimental value of $1{\cdot}55 \pm 0{\cdot}2$ at 40 MEV and $3{\cdot}0 \pm 0{\cdot}1$ at 90 MEV, we see that the compact cut-off potentials predict angular distributions that are too flat at 40 MEV and distributions that show too much angular variation at 90 MEV, while the long-tailed potentials show fair agreement with experiment at both energies.

The only partial waves contributing appreciably to the cross section are the S and D waves below 90 MEV; consequently the angular distribution can be expanded in terms of Legendre polynomials P_0, P_2 and P_4. The coefficient of P_0 is identical with the total cross section, that of P_2 arises primarily from the interference between the S and D states, and that of P_4 arises from interference among the D states. These coefficients allow a rapid comparison of theory and experiment (see table 4 for explicit comparison of predicted scattering and experimental results at 40 MEV and 90 MEV).

Yukawa and exponential potentials (see figs. 4 and 9) show a somewhat flatter cross section, in the neighbourhood of 90°, than the experiments. This is manifested by a P_4 coefficient that is two or three times larger than the experimental value.

We shall not try to obtain the best compromise between the total cross section and the angular distribution but rather shall try to make the angular distribution fit as well as possible. The reason for this choice is twofold. First, the total cross section in n–p scattering is more susceptible to systematic errors than the

Table 4

Model				90 MeV							40 MeV		
Exchange dependence	Range [10^{-13} cm]	Radial dependence	γ	σ (10^{-26} cm^2)	a_1	a_2	a_3	a_4	$\frac{\sigma(180)}{\sigma(90)}$	$\frac{\sigma(0)}{\sigma(90)}$	σ (10^{-26} cm^2)	a_2	$\frac{\sigma(180)}{\sigma(90)}$
$(1+P_M)/2$	1·18	Y	0	9·0	0	0·77	0	0·39	3·25	3·25	23·1	0·15	1·26
$(1+P_M)/2$	1·18	Y	5·6	9·9	0	0·75	0	0·04	2·91	2·91	—	—	—
$(1+P_M)/2$	1·35	Y	0	9·3	0	0·98	0	0·57	4·6	4·6	22·9	0·21	1·45
$(1+P_M)/2$	1·35	Y	1·9	10·2	0	0·78	0	0·14	3·20	3·20	23·1	0·24	1·46
$(0·37+0·63P_M)S_{12}$*	1·35	Y	1·9	10·7	−0·20	0·70	0·12	0·12	3·04	2·95	—	—	—
$(0·24+0·76P_M)S_{12}$*	1·35	Y	1·9	12·0	−0·35	0·66	0·24	0·06	2·61	2·46	—	—	—
$0·45+0·55P_M$	1·35	Y	1·9	10·3	−0·16	0·78	−0·02	0·15	3·52	2·84	—	—	—
$0·4+0·6P_M$	1·35	Y	1·9	10·4	−0·32	0·77	−0·05	0·16	3·78	2·48	—	—	—
$(1+P_M)/2$	0·7	E	0	7·9	0	0·99	0	0·39	4·00	4·00	21·5	0·17	1·30
$0·45+0·55P_M$	0·7	E	0	7·9	−0·10	0·99	−0·07	0·39	4·33	3·69	—	—	—
$0·4+0·6P_M$	0·7	E	0	8·0	−0·20	1·00	−0·16	0·41	4·68	3·39	—	—	—
$(1+P_M)/2$	0·75	E	1·8	8·7	0	0·92	0	0·03	3·8	3·8	21·7	0·18	1·33
$-(\sigma_1 . \sigma_2)(\tau_1 . \tau_2)/3$	2·0	S	0	7·1	−0·86	1·13	−0·34	0·12	9·57	1·59	21·3	0·11	1·42
$-(\sigma_1 . \sigma_2)(\tau_1 . \tau_2)/3$	1·8	S	0	1·4	−0·61	0·63	−0·19	0·05	3·50	1·30	22·2	0·04	1·14
Experimental value				7·9±1·0	−0·14±0·10	0·68±0·10	0·02±0·10	0·11±0·10	3·0±0·2	3·0±1·0	19·4±2·0	0·26±0·10	1·55±0·20

* The exchange dependence for the central force is $(1+P_M)/2$.

High energy scattering behaviour of various models. In the above, range (R) is the same in singlet and triplet states. For all cases where $\gamma \neq 0$, γ is adjusted to a quadrupole moment of $2·73 \times 10^{-27}$ cm^2. σ is the total cross section, the differential cross section being $4\pi\sigma(\theta) = \sigma . \Sigma a_n P_n(\theta)$, where $a_0 = 1$.

measurements of the relative angular distribution at a particular energy, and, secondly, relativistic corrections are probably more important for the S state interaction, where the kinetic energy of the nucleons inside the range of the potential becomes considerably higher than that in the external region. It might further be mentioned that a slight decrease in the S phase shift leads to an almost uniform decrease in the angular distribution. This results because the scattering amplitude can be analysed into a real and imaginary part, so that the expression for the total cross section becomes

$$\sigma(\theta)=\frac{1}{k^2}\{[\Sigma_L \sin^2\delta_L P_L(\theta)]^2+[\Sigma_L \sin\delta_L \cos\delta_L P_L(\theta)]^2\}. \quad \ldots\ldots(7)$$

The main angular variation comes from the second term, in which the combination $\sin\delta_0 \cos\delta_0$ occurs. For S phase shifts between 30° and 60° the total variation

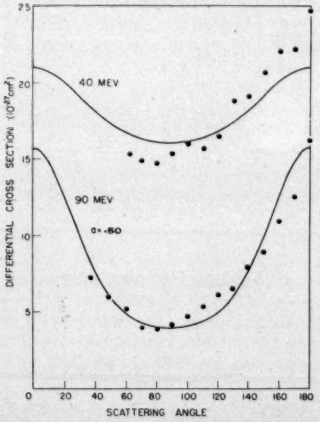


Fig. 9. n–p scattering at 40 and 90 MEV from an exponential potential, $\exp(-r/R)$, ($R=0{\cdot}7\times10^{-13}$ cm for both singlet and triplet states) without tensor forces. The exchange dependence is $\frac{1}{2}(1+P_M)$. The total cross sections for this potential are $0{\cdot}217\times10^{-24}$ cm^2 and $0{\cdot}079\times10^{-24}$ cm^2 at 40 and 90 MEV, respectively. The experimental points have been normalized to the theoretical angular distribution.

of this combination is only 13%. In the first term, which gives an almost uniform contribution to the cross section (due to the smallness of the higher phase shifts), the variation of $\sin^2\delta_0$ between 30° and 60° is 300%.

(iii) *Tensor Forces*

The existence of the deuteron quadrupole moment requires the inclusion of a tensor potential in the neutron–proton interaction. If the same radial form is assumed for the central and tensor potential, then the depths of both the central and tensor potential, as well as their joint range, are determined by this additional requirement. It would be possible to determine the relative central and tensor ranges separately if it could be assumed that the difference between the deuteron magnetic moment and the sum of the neutron and proton magnetic moments gave

a measure of the D state admixture. Unfortunately, this quantity is especially susceptible to relativistic correction and forms an unreliable restriction (Sachs 1947, Primakoff 1947, Breit and Bloch 1947).

As in the central force case, an effective range theory can be given (Christian 1949), the only difference being in the relation of the zero-energy solution to the effective range. The ground state is now the coupled state $^3S + {}^3D$, and thus the radial equations are coupled second order differential equations. Calling the S state component $u(r)$ and the D state component $w(r)$, the expression for the effective range becomes

$$r_0 = 2\int_0^\infty \left[\left(\frac{r}{a}+1\right)^2 - u^2(r) - w^2(r)\right] dr, \qquad \ldots\ldots(8)$$

where $u(r)$ approaches $r/a+1$ asymptotically, and $w(r)$ approaches zero. The parametric ranges may now be calculated as before, with the further requirement that the bound deuteron exhibit the correct quadrupole moment.

In the case of the square well this calculation leads to a parametric range of only $2{\cdot}3 \times 10^{-13}$ cm, and the entire binding in the deuteron results from the tensor force, the central force vanishing completely. For the exponential potential the range required is $0{\cdot}75 \times 10^{-13}$ cm, and triplet central depth being 69 Mev and the tensor depth being 128 Mev. The Yukawa potential requires a range of $1{\cdot}35 \times 10^{-13}$ cm, a triplet central potential depth of 25·3 Mev and a tensor depth of 48·2 Mev.

The above calculations show that the introduction of the tensor force causes relatively small changes in the low-energy scattering properties even though the central depth may change by a large amount. Such behaviour can be expected for more general potential shapes with ranges short compared with the deuteron radius, since the S wave component is determined primarily from the boundary condition at the origin and the asymptotic deuteron behaviour (Hefner and Peierls 1942).

We can put these arguments on a more quantitative basis by the consideration of an 'equivalent central potential' '$V(r)$'. Writing the potential in the form $V(r) = V_c(r) + \gamma S_{12} V_t(r)$ where $V_c(r)$ and $V_t(r)$ are the central and tensor potentials, even parity potentials, respectively, and S_{12} is the tensor operator,

$$3(\boldsymbol{\sigma}_1 \cdot \mathbf{r}_{12})(\boldsymbol{\sigma}_2 \cdot \mathbf{r}_{12})/r_{12}^2 - \boldsymbol{\sigma}_1 \cdot \boldsymbol{\sigma}_2,$$

the equivalent central potential for the S wave is

$$'V(r)' = V_c(r) + 2^{3/2}\gamma V_t(r) R(r), \qquad \ldots\ldots(9)$$

where $R(r)$ is the ratio of the D wave component to the S wave component. $R(r)$ will be, in general, a slowly varying function of the energy (at least in the region where the potential is large). Its form may be estimated from considerations of the bound state solutions. It is found that $R(r)$ is zero at the origin, increases to a maximum value (about 0·2 to 0·3) somewhere between the maximum value of the S wave radial function and the tensor force range, and decreases asymptotically to a small value (somewhat below 0·1). The ratio of the equivalent potential '$V(r)$' to the central potential $V_c(r)$ (the latter adjusted to give the correct binding by itself) is less than unity in the neighbourhood of the origin. The ratio increases to approximately unity in the neighbourhood of the range and decreases asymptotically if the tensor range is the same as the central range. If the tensor range is longer than the central range there would be no decrease in the ratio of the 'equivalent central range' to the central range.

This can be further illustrated in terms of the WKB approximation (for details see Christian and Hart 1950). In this approximation $R(r)$ is independent of energy and decreases asymptotically to zero. The equivalent potential in this approximation is

$$'V(r)' = V_c(r) - \gamma V_t(r) - \frac{3}{r^2} + \left[\left\{\gamma V_t(r) + \frac{3}{r^2}\right\}^2 + 8\{\gamma V_t(r)\}^2\right]^{1/2}. \quad \ldots\ldots(10)$$

Since the centrifugal potential is usually large compared with the tensor potential at all distances, this may be approximated by

$$'V(r)' = V_c(r) + \tfrac{4}{3}[r\gamma V_t(r)]^2, \quad \ldots\ldots(11)$$

which is clearly in agreement with the preceding remarks.

Turning now to the high energy scattering, we can expect that the scattering in S and D partial waves will contribute almost the whole of the cross section, since the range of forces has changed only slightly from that in the central force model.

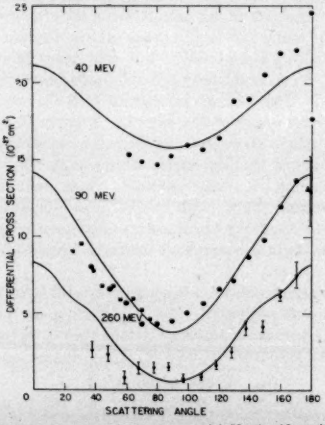


Fig. 10. n–p scattering from an exponential tensor potential, $V = (-69 \text{ MEV})(1 + 1{\cdot}84 S_{12}) \exp(-r/R)$, with $R = 0{\cdot}75 \times 10^{-13}$ cm. The total cross sections are $0{\cdot}217 \times 10^{-24}$ cm² and $0{\cdot}087 \times 10^{-24}$ cm² at 40 and 90 MEV, respectively. The experimental points have been normalized to fit the theoretical angular distribution.

For a somewhat more detailed comparison we could again use the WKB approximation to approximate the 'equivalent central potentials' for each of the states L and J. In this approximation we neglect the asymptotic amplitude of the coupled mode in the evaluation of the phase shifts, and there will be no difference between the states of different magnetic quantum number m_s. The WKB approximation yields results, however, that agree within 10 to 20 per cent with those of more accurate calculations in the case of the Yukawa potential, so that it appears reasonable to use it for a qualitative understanding of the effect of tensor forces. The results for the various 'equivalent central potentials' in the D states indicate that the radial form of the potential is not radically changed even though the magnitude of the various states of differing J may be changed.

Utilizing this comparison between the tensor and central force models, we can conclude that the potential must be long-tailed in order to maintain the relatively large scattering in the D state at 40 Mev. Cut-off potentials (e.g. square and Gauss), where the tensor force has the same or nearly the same range, are, therefore unsuitable.

The results of using the Yukawa and exponential radial forms (figs. 10, 11 and 12) indicate that the addition of the tensor force indeed causes little change in the

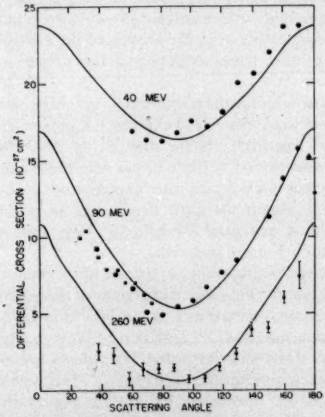


Fig. 11. n–p scattering from a Yukawa tensor potential,

$$V=(-25{\cdot}3 \text{ Mev})(1+1{\cdot}91S_{12}) \exp(-r/R)/(r/R),$$

with $R=1{\cdot}35\times10^{-13}$ cm. The total cross sections are $0{\cdot}231\times10^{-24}$ cm², $0{\cdot}098\times10^{-24}$ cm², and $0{\cdot}037\times10^{-24}$ cm² at 40, 90 and 280 Mev, respectively. The experimental points have been normalized to fit the theoretical angular distribution.

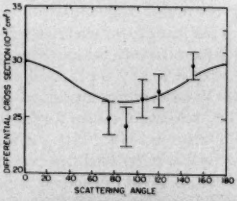


Fig. 12. The 27·2 Mev n–p angular distribution (Brolley, Coon and Fowler 1951). The solid line is the prediction of the Yukawa tensor model.

scattering. The best fits of the angular distribution require a slightly longer parametric range, in agreement with the increase required by the low-energy scattering. It can be noted that there is a considerably better fit in the detailed shape with the inclusion of the tensor force. With the tensor force the calculated angular distribution is V-shaped, whereas with central forces alone the distribution

is more U-shaped. A detailed comparison using the Legendre coefficients shows that the $P_4(\theta)$ component is reduced in the tensor force model compared with the central force model, more in agreement with experiment (see table 4). Some understanding of this effect may be gained by considering the tensor forces in Born approximation. As we have seen, the tensor forces are far from weak and so the Born approximation cannot be expected to give more than the general trend. The characteristic peaking of the Born approximation cross section around 45° and 135° (the exact angle depending upon the model, range and energy, with a maximum occurring roughly where $2kR\sin\frac{1}{2}\theta\sim1$) is the type of correction needed to explain the discrepancy between the shapes of the central and tensor angular distributions. This, then, gives evidence of the presence of tensor forces in scattering.

The same situation holds for the tensor model as for the central model regarding the intercomparison of the Yukawa and exponential potentials (cf. figs. 10 and 11). That is, the Yukawa potential fits the angular distributions noticeably better; however, the total cross section is 20 to 30 per cent higher than the experimental value. The total cross section for the exponential potential is only 10 to 15 per cent too high. Again the same arguments as in the central force case require the long-tailed potential to have a deep hole at the origin and, consequently, too much S wave scattering.

The principal change arising from an increase of the tensor range relative to the central force range is that the long-tailed character of the potential is increased and the strength of the tensor interaction must be less in order to obtain the correct value for the quadrupole moment. There is very little change in the high-energy scattering in the case of the Yukawa radial dependence when the tensor range is increased if the low-energy parameters are fitted. The characteristic tensor peaks become more pronounced and move to smaller angles when the tensor range becomes very long. This limits the parametric tensor range with the Yukawa potential to $2{\cdot}5$–$3{\cdot}0\times10^{-13}$.

Pease and Feshbach (1950) have investigated the binding energy of ^{3}H using Yukawa wells for both the central and tensor forces. They find that in order to get the correct binding energy, and if the central range is taken to be the same as that required for p–p scattering, the tensor range must be increased relative to the central range by a factor of 1·4. This leads to an effective range for the triplet n–p state of $1{\cdot}74\times10^{-13}$ cm, in good agreement with the experimental value. The scattering cross section calculated at 40 and 90 Mev is practically the same as that calculated for the $1{\cdot}35\times10^{-13}$ cm range Yukawa tensor model shown in fig. 11.

Further evidence for the presence of tensor forces can be found by examining the polarization effects (Wolfenstein 1949, 1951). Using the Yukawa even parity tensor model Swanson (1951) has calculated the expected azimuthal asymmetry in a double scattering. Wouters (1951) produced a partially polarized neutron beam by a (p, n) reaction of LiD and detected a maximum azimuthal asymmetry of the order of 8% on the subsequent n–p scattering. This preliminary figure should be compared with a predicted 12%. The presence of tensor forces in high-energy scattering is thus definitely established since central forces alone lead to no asymmetry.

(iv) *Odd Parity Potentials*

The experiments at 90 Mev actually indicate a slight asymmetry around 90°. We can calculate the effect of adding a slight amount of odd parity force to the central force model by using the common factor $(1-a+aP_M)$ in the singlet and

triplet central potentials. P_M is the space exchange operator, so that with $a=\frac{1}{2}$ the odd state potentials vanish, $a=1$ corresponds to exchange forces and $a=0$ corresponds to ordinary forces. The results are shown in fig. 13 for the Yukawa

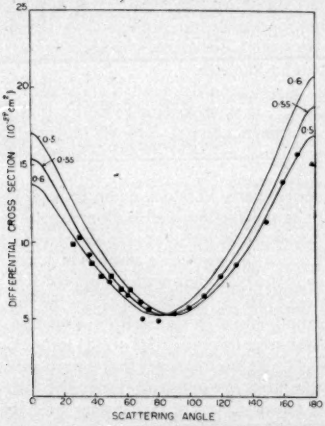


Fig. 13. n–p scattering at 90 MeV from the Yukawa tensor potential ($R=1{\cdot}35\times10^{-13}$ cm) showing the effects of increasing the amount of exchange force by the use of the exchange factor $(1-a+aP_M)$ which multiplies the even state potential. The curves are labelled by the value of a.

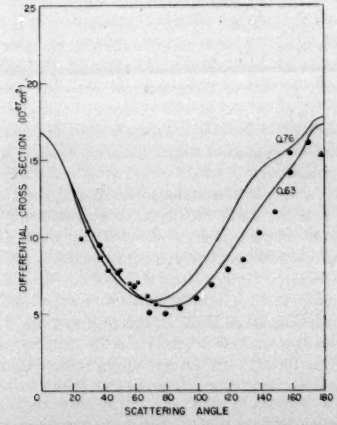


Fig. 14. n–p scattering at 90 MeV from the Yukawa tensor potential ($R=1{\cdot}35\times10^{-13}$ cm) showing the effect of increasing the amount of odd tensor force with no odd central forces present. The curves are labelled by the value of a used in the exchange factor $(1-a+aP_M)$, which multiplies the even state tensor potential.

potential and lead to the limitation $a = 0{\cdot}55 \pm 0{\cdot}05$. More generally, allowing for a different exchange character in even and odd states, the depths of the odd parity potential V_{odd} must satisfy the approximate relation

$$0 < \tfrac{1}{4}({}^1V_{\text{odd}}) + \tfrac{3}{4}({}^3V_{\text{odd}}) < -\tfrac{1}{3}({}^3V_{\text{even}}).$$

This specification is not completely independent of the radial form and range of the potential in the odd states, but holds in particular for the long-tailed potentials with equal ranges in even and odd states.

The limitation on the odd central potentials arises primarily from the increase in asymmetry about 90°. It is important to notice that since the odd central scattering amplitude is antisymmetrical about 90°, there is hardly any shift in the minimum and, consequently, no better fit of the angular distribution than with only even forces.

The effect of introducing an odd tensor potential is indicated in fig. 14. The improvement in the fit of the angular distribution is quite noticeable and results from the fact that the tensor forces introduce the associated Legendre polynomials in the scattering, and hence the scattering amplitude is not necessarily antisymmetric and can thus shift the minimum without appreciably altering the value of the cross section at 0° and 180°. The limitation on the magnitude of the tensor arises primarily from the increase in the total cross section.

Since there are no repulsive forces in a $\frac{1}{2}(1 + P_M)$ interaction, it is impossible to account for the saturation, and it is necessary that the repulsive forces be more than one-third as strong as the attractive forces (see especially Rosenfeld 1948, chap. 11, Gerjuoy 1950). Odd tensor forces of either sign probably do not meet the saturation requirements either.

§ 4. PROTON–PROTON SCATTERING

(i) *Qualitative Discussion*

The success that was obtained in the n–p system would seem to be sufficient grounds for expecting that p–p scattering could also be interpreted in terms of a static potential. As we have seen in the introduction, the predictions based upon charge symmetry were in violent disagreement with the data.

Let us consider in more detail the reasons for the breakdown of charge independence in the hope of gaining an insight into p–p scattering. We found that the n–p potential must predict very little scattering in the odd states, and if there were actually no interaction whatsoever in the odd states then, on the basis of charge independence, the only scattering in the p–p system would be the singlet states. This is because the Pauli exclusion principle requires in the p–p system that the singlet states be of even parity and the triplet states of odd parity. Thus in the absence of any odd parity scattering the only states perturbed would be the ^1S, ^1D, ^1G, etc. From the study of the n–p system we found that at energies below 100 Mev the only nuclear scattering occurs from states with $L \lesssim 2$. Of course, coulomb scattering takes place in the p–p system, but this is dominant 32 Mev only for angles less than 20°. Between 20° and 50° the coulomb-nuclear interference will be important. In the remaining region around 90° the coulomb scattering amplitude is so small that the scattering is virtually the same as simple nuclear scattering.

It was observed immediately, as pointed out in the experimental papers (Cork *et al.* 1950, Panofsky and Fillmore 1950) that the 32 Mev data was in good agreement

with that predicted by S wave scattering alone. This is in definite disagreement with the scattering predicted by any of the radial forms considered previously. The disagreement results because the low-energy data require a radius so large that the D wave predicted at this energy, even for the most cut-off potential (i.e. a square well), leads to a cross section having a dip in passing from 50° to 90°. Thus the predicted cross section is similar in appearance in this angular range to that actually found in the n–p experiments (see fig. 1). Since the square-well potential predicts too large a D phase shift, such a behaviour may be expected generally for all monotone potentials. (It would, of course, be possible to choose a radius sufficiently small that the D wave is not perturbed at 32 Mev, but this would lead to a range much too short to fit the low-energy data.)

Actually very little was determined from the n–p data regarding the singlet radial potential shape, because of the possibility of its scattering cross section being masked by the triplet state which has the higher statistical weight. We might, therefore, attempt to find a more complicated radial dependence, such as a square well with a repulsive lip, which could lead to a negligible D phase shift at 32 Mev. The inherent difficulty of fitting the 340 Mev data with such a model rules against it.

The 340 Mev scattering is even more strikingly anomalous. The experiments indicate a nearly spherical symmetric distribution from 15° to 90° (centre-of-mass angles) of nearly twice the maximum possible for S wave scattering alone. (The maximum S wave scattering is $(\lambda/2\pi)^2$/sterad, which at 350 Mev is $2{\cdot}5 \times 10^{-27}$ cm^2/sterad, while the experimental value is close to $4{\cdot}0 \times 10^{-27}$ cm^2/sterad.) Any model with central forces alone leads to a cross section that is sharply peaked in the forward direction. The experimental results are superficially similar to classical hard sphere scattering, but the radius of the hard sphere to give a differential cross section of $4{\cdot}0 \times 10^{-27}$ cm^2/sterad is only $1{\cdot}0 \times 10^{-13}$ cm, which is only twice the wavelength of the proton, so that the energy is far too low to show this type of scattering. Further experimental results show that the cross section remains isotropic (over the angular range measured) down to 100 Mev, with little change in the absolute magnitude of the cross section. Thus an attempt to fit the data by such a model is out of the question. Since the n–p scattering at 90 Mev and 260 Mev is in good agreement with the predictions of a static potential, it is difficult to accept the discrepancy between the predictions of a static potential and the experimental result as a relativistic effect.

We could attempt as an alternative to predict the p–p scattering by requiring that the p–p forces make up for the n–p forces in explaining the act of saturation. The n–p experiments rule out anything like the magnitude required to explain saturation, and the low-energy p–p experiments show that the singlet force is attractive. For the p–p forces to give saturation there must be a strong repulsive force in the triplet states. Since the scattering amplitude from a central triplet potential is antisymmetric about 90°, and there is no interference with the singlet state, there must be an addition to the p–p cross section increasing on either side of 90°. This leads then to an even worse disagreement with the data than the singlet state alone, and such a model can be ruled out.

Thus both the hypothesis of charge independence and the possibility of central triplet forces appear to be ruled out by both the 32 Mev data. The remaining alternative, within the framework of the potential picture, is the possibility that tensor forces might compensate in such a manner at both energies so as to lead to a spherical symmetric cross section. A favourable result is predicted by the use of

the Born approximation to compute the scattering. (The Born approximation becomes more nearly valid for P and higher waves since the centrifugal barrier reduces the effect of the nuclear potential to a small perturbation.) The tensor cross section can be peaked at 90° at any particular energy with any potential model by a proper choice of range. This can add to the singlet cross section which dips at 90° leading to an almost flat nuclear cross section. In the next section, devoted to the p–p scattering calculations, we shall find that it is indeed possible to reconcile all the p–p data with a static potential using such a model.

(ii) *Proton–Proton Calculations*

The cross sections at 32 Mev and 340 Mev predicted by various singlet potentials adjusted to fit the low-energy data are shown in fig. 15 (see table 2 for range and depth). The Yukawa potential predicts at 32 Mev not only the largest cross section but also the most angular variation in the 20° to 90° range. This can be understood because the deep hole at the origin leads to high S phase shifts and the long tail leads to large D phase shifts. (The ^{1}S phase shifts involved are 41·99°, 47·54° and 51·15°, the ^{1}D phase shifts are 0·77°, 1·20° and 1·40° for the square, exponential and Yukawa potentials respectively.)

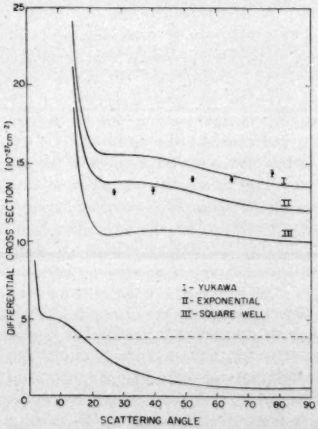


Fig. 15. Singlet p–p scattering at 32 and 340 Mev from various potentials adjusted to fit the low-energy scattering. The scattering from all the potentials at 340 Mev is similar and disagrees violently with the experimental data (represented by the broken line).

None of these predictions are in qualitative agreement with the experimental results. At 340 Mev the disagreement is especially large. The predicted cross section is peaked in the forward direction and falls to a small fraction of the experimental value at 90°.

The addition of central odd parity forces does not change the 90° cross section, so that if we wished to fit the 32 Mev data without regarding the 340 Mev data, the Yukawa potential must be used (since it comes closest to fitting the 90° point).

However, we have already seen in fig. 5 that the effect of adding attractive and repulsive central odd forces does not give any better agreement than the central even forces alone. This is because in the 50° to 90° region, where coulomb effects are unimportant, the P state cross section has the form $\cos^2\theta$ and is additive, and can only increase the angular variation of the cross section. Thus it is shown conclusively that the 32 Mev data alone require other than central forces.

Likewise the 340 Mev data strongly indicate the presence of non-central forces. (It does not appear possible to construct a singlet potential, even if the radial dependence is not monotonic, that can give a spherical symmetric cross section at 340 Mev primarily due to the limitations imposed by the effective range. The small effective range results in scattering only from states with angular momentum 0, 2

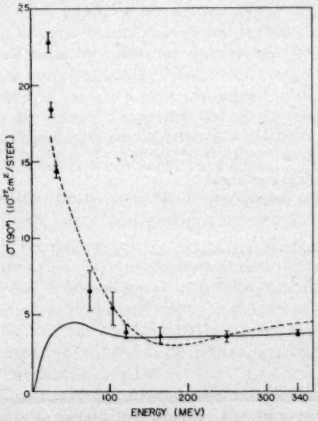


Fig. 16. The p–p scattering cross section at 90 degrees as a function of energy. The solid curve represents the difference between the experimental cross section and that predicted by a square-well potential in the singlet state. The broken line shows the predictions of the repulsive core model (see fig. 21).

and 4, and consequently there are only three phase shifts. Even if these are chosen arbitrarily they cannot be made to reproduce the experimental cross section.)

The possibility of masking the singlet behaviour, mentioned in the previous section, by triplet tensor scattering requires that the magnitude of the singlet cross section be smaller than the experimental value. Thus at 32 Mev only a shallow (hence cut-off) potential such as the square or Gauss potentials can be used. With the choice of the square-well potential in the singlet state the magnitude of the triplet cross section at any angle (and energy) is calculable simply by subtracting the predicted singlet cross section from the measured value. (Actually it makes little difference whether the square or Gauss potential is used.) The 90° value of the triplet cross section so derived gives the nuclear contribution directly since the interference terms with the coulomb scattering drop out at this angle. Fig. 16

shows the experimental value of the 90° cross section and also the result of subtracting the singlet cross section. Fig. 17 shows the predictions of the Yukawa and the singular tensor potential, each adjusted in range and depth, so that they give about the same quality of fit at 32 MeV. Thus on the assumption that the singlet scattering is given by a square-well potential, the singular tensor potential is required to fit the high-energy data. The 32 MeV and higher energy fits with this combination are shown in fig. 18.

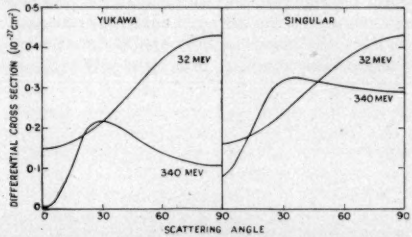


Fig. 17. The odd state tensor scattering contributions at 32 and 340 MeV. The magnitude has been adjusted so that a good fit of the 32 MeV data is obtained when this potential is added to the square-well singlet potential.

Note: The ordinate scale should be increased by a factor of 10.

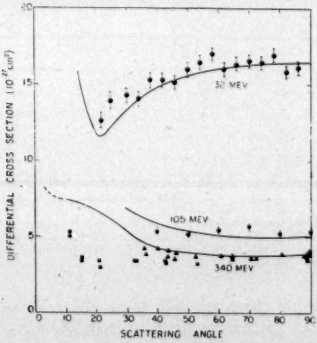


Fig. 18. p–p scattering at various energies from the combination of a square well (parametric range $2{\cdot}615\times10^{-13}$ cm and $-13{\cdot}273$ MeV deep) in the singlet state and a singular potential in the triplet state $[V=(-15{\cdot}25\text{ MeV})S_{12}\exp(-r/R)/(r/R)^4$, with $R=1{\cdot}6\times10^{-13}$ cm.]

The fit of the 32 and 105 MeV data is excellent, but the theoretical predictions at 340 MeV are considerably higher than the experimental results in the 10 to 25° angular range. It is not clear how to overcome this difficulty as most of the cross section in this angular range is the result of scattering in the singlet state and the singlet scattering at this energy depends very little upon the explicit radial form chosen. It is consequently desirable that the experiments in the 100 to 200 MeV range be extended to smaller angles to see if the discrepancy is present at these intermediate energies also.

(The tensor force calculations have only been carried out in Born approximation. The comparison with cross sections calculated more accurately shows that there may be errors of the order of 20 per cent. The exact calculations in general predict a somewhat more spherical symmetric cross section than the Born approximation. A refinement of the Born approximation has been developed (Christian and Noyes 1950) to determine the tensor–coulomb interference terms, which vanish in Born approximation. It is found that there is no difference in sign or magnitude of the tensor–coulomb interference terms for either sign tensor potential, so that from p–p data alone it is impossible to tell the sign of the tensor potential. This may be contrasted with the central coulomb interference term where it is possible from the sign of the interference term to tell the sign of the central potential. Actually, from a computational standpoint the singular tensor potential cannot have the negative sign because the singularity introduced at the origin in the $^3P_0$ radial equation is too strong to permit a solution (other than in Born approximation). This is probably only a formal difficulty since in Born approximation the scattering produced by a singular potential can be reproduced closely below 340 MEV by a sum of two Yukawa potentials, one of parametric range 1×10^{-13} cm and the other of parametric range $0{\cdot}4 \times 10^{-13}$ cm. The possibility of requiring the tensor depth to be so strong that a $^3P_0$ bound state results appears unlikely. Further calculations in other than Born approximations are urgently needed to settle some of the outstanding questions.)

§ 5. THE CHARGE-INDEPENDENT HYPOTHESIS

We have found that the differences in n–p and p–p scattering made it necessary to determine their interactions independently and that the potential models were quite different. This is in disagreement with the hypothesis of charge independence. The hypothesis is in turn based upon the concept of protons and neutrons as different charge states of one fundamental particle—the nucleon. It is consequently of major importance, from a theoretical standpoint, to find out if a charge-independent formulation is possible.

We have already found a charge-independent model capable of explaining the data. The problem is to determine whether the lack of uniqueness in both the n–p and p–p models is sufficient to allow us to find a charge-independent model. It is thus clear that we will not be able to prove the hypothesis even if a charge-independent formulation is possible. On the other hand, it would be highly unlikely if it were shown experimentally that the neutron–proton scattering is exactly symmetrical about 90°, and therefore contains no scattering in any kind of odd parity state. It is just the small asymmetry suggested by the present 90 MEV n–p scattering that might allow the hypothesis to be retained. The most obvious difference in the two interactions is in the triplet odd states. The n–p system can tolerate only a small fraction of its total scattering to be in odd states, while at high energy the p–p system requires the major fraction of its scattering to be in the odd triplet states. Because of the exclusion principle there are two differences between n–p and p–p scattering; first, the even triplet and the odd singlet states are missing in p–p scattering and, secondly, the value of the differential cross section from a particular state is four times larger in the p–p system than in the n–p system for the same potential interaction. Noting the second point, we may calculate the contribution to the n–p total cross section upon the introduction of the odd state tensor potential necessary in the p–p system. We have found

(fig. 17) that the tensor contribution is an almost uniform cross section of 4×10^{-27} cm^2/sterad (as a function of angle and energy about 32 Mev). This will of course not be a uniform addition to the n–p cross section because of the interference with the even triplet states. The contribution to the n–p total cross section will be $\frac{1}{4} \times 4 \times (4\pi) \times 10^{-27}$ cm^2, since the interference terms cancel upon integration. The measured n–p cross section at 90 Mev is 75×10^{-27} cm^2, so that this would be a relatively small fraction of the total scattering. However, the predicted total cross sections from even states alone is somewhat too high already (87×10^{-27} cm^2 for the exponential interaction and 102×10^{-27} cm^2 for the Yukawa interaction), so that an additional 12×10^{-27} cm^2 is hard to tolerate. The effect on the n–p angular distribution caused by the addition of the singular tensor force is shown in fig. 19. The theoretical curve is shown with a singular potential

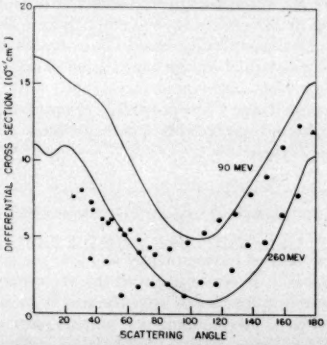


Fig. 19. n–p scattering on the basis of charge independence, using the same potential as that shown in fig. 18 in the even singlet and odd triplet states. The potential used in the even singlet states is $V=(-25{\cdot}3 \text{ Mev})(1+1{\cdot}91S_{12}) \exp(-r/R)/(r/R)$, with $R=1{\cdot}35 \times 10^{-13}$ cm. The potential is zero in the odd singlet states.

having a depth of $-15{\cdot}25$ Mev and range of $1{\cdot}6 \times 10^{-13}$ cm. The opposite sign for the potential would in Born approximation just change $\sigma(\theta)$ into $\sigma(\pi-\theta)$, and would show slightly better agreement with the experimental results. It can be estimated that the strength must be 8 Mev or less for a good fit. It appears that although there is some hint of odd tensor forces in n–p scattering, the magnitude required by charge independence is about twice too large.

It has been suggested by Jastrow (1951) that at short distances the central potential becomes strongly repulsive. This repulsion occurs around $0{\cdot}6 \times 10^{-13}$ cm in singlet states and around $0{\cdot}2 \times 10^{-13}$ cm in triplet states. The interaction can be represented by a hard sphere of the proper radius and a surrounding attractive potential having an exponential radial dependence of such depth and range that the low-energy parameters are given correctly. The interaction in the singlet states takes the form

$$V=\infty, \qquad r<r_0,$$
$$V=\tfrac{1}{2}(1-P_\sigma P_r)V_0 \exp[(-r-r_0)/r_s], \qquad r>r_0,$$

with $r_0=0{\cdot}6 \times 10^{-13}$ cm, $V_0=375$ Mev and $r_s=0{\cdot}4 \times 10^{-13}$ cm. The theoretical p–p cross section at 32 and 340 Mev predicted by such a potential is shown in fig. 20.

The cross section at 32 MeV is almost the same as that predicted for a square-well potential, but the cross section at 340 MeV has a minimum around 40° and rises to a value of almost 3×10^{-27} cm²/sterad at 90°. If the triplet tensor scattering shows a maximum in the cross section at about the same angle that the singlet scattering shows a minimum it is possible for the sum of the two types of scattering to lead to a more nearly isotropic cross section than either one alone. By referring to figs. 17 and 20 we see that it is possible to adjust the depths of the Yukawa tensor

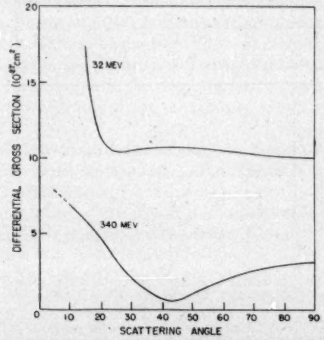


Fig. 20. p–p scattering predicted by the use of the repulsive core potential in the singlet states ($V=\infty$ when $r<r_0$, $V=(-375$ MeV) exp $[(-r-r_0)/r_t]$ when $r>r_0$, where $r_0=0{\cdot}60\times10^{-13}$ cm and $r_t\times0{\cdot}40=10^{-13}$ cm.

potential so that the sum of the tensor scattering and the scattering from the repulsive core singlet potential is nearly isotropic except at small angles. A better fit of both the 32 MeV and 340 MeV data is, however, attained by the use of an exponential radial form for the tensor potential. The result for the interaction in triplet states is then

$$V=\{0{\cdot}5-0{\cdot}5P_\sigma P_\tau+\gamma(0{\cdot}3-0{\cdot}7P_\sigma P_\tau)S_{12}\}V_0\exp(-r/r_t),$$

with $V_0=-69$ MeV, $\gamma=1{\cdot}84$, and $r_t=0{\cdot}75\times10^{-13}$ cm. (The repulsive core has been left out.) The p–p scattering predictions at 32, 105 and 340 MeV are shown in fig. 21. The agreement with the data is excellent at 340 MeV except in the 10° to 25° angular range, where the predictions are considerably higher than the experimental results. The calculated p–p cross section at 90° as a function of energy is shown by a broken line in fig. 16. The asymmetry introduced by the addition of the odd state triplet potential in n–p scattering at 260 MeV is shown in fig. 22. n–p cross sections at 90 MeV have not as yet been calculated, but it can be expected that the asymmetry will not be quite as marked as with the singular tensor model but will show considerably more asymmetry than is indicated by the experiments. The theoretical n–p cross section at 40 MeV will probably be flatter than is shown by experiment. (A preference could probably be established for either the repulsive core or for the singular tensor model by extending the 105 MeV p–p data to smaller angles, as can be seen from figs. 18 and 20).

Jastrow has noted that the short range repulsion will lead to a type of saturation in heavy nuclei analogous to that existing in classical liquids. Since the radius of repulsion is at least a factor of four smaller than the inter-nucleonic spacing it

appears questionable that reasonable values of the binding energies and densities will result with this form of interaction. It is important to note, however, that the effect of the exclusion principle may lead to spacings larger than the radius of repulsion in analogy with the situation in liquid helium, where the particles are coupled by an interaction similar to the above except for scale. Detailed calculations have not as yet been carried out concerning nuclear saturation.

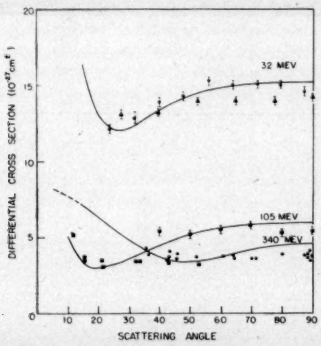


Fig. 21. p–p scattering predicted by the addition of a tensor potential in the triplet state to the repulsive core potential shown in fig. 20. In odd triplet states the potential is $V=(27{\cdot}6$ MeV) $\exp(-r/r_t)$, where $r_t=0{\cdot}75\times10^{-13}$ cm.

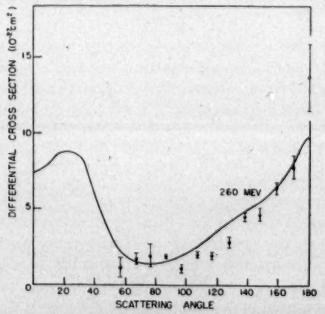


Fig. 22. n–p scattering at 260 MeV on the basis of charge independence using the repulsive core potential in even singlet states (see fig. 20), and the tensor potential of fig. 21 in odd triplet states. The potential in even triplet states is $V=(-69$ MeV$)(1+1{\cdot}84\,S_{12})\exp(-r/r_t)$ with $r_t=0{\cdot}75\times10^{-13}$ cm. The potential is zero in the odd singlet states.

Case and Pais (1950) have presented still another manner in which to explain the p–p scattering data that may be incorporated into a charge-independent interaction. They use a velocity-dependent spin–orbit force in the triplet odd parity states. Calculations carried out in Born approximation indicate that, as with the tensor potential, a strongly singular potential is required. There interaction in the even parity states is the same as that used for the singular tensor charge-independent model. It is probable that the agreement with the n–p data is no

better than with the singular tensor potential model, even though detailed calculations of the even–odd parity interference terms in n–p scattering have not been carried out.

A preference for any of these models at the present time would be premature, since better theoretical calculations are needed. There are some discrepancies between theory and experiment that are common to any of these models. First, the n–p total cross section predicted is 20 to 30 per cent higher than the experimental value at 40 and 90 Mev. Second, the predicted n–p angular distributions tend not to show sufficient angular variation at 28 and 40 Mev. Third, the flatness of the p–p cross section down to angles as small as 15° is not predicted by any of these models. In addition, the introduction of the triplet odd state potential (i.e. the one necessary for p–p scattering) into the n–p interaction appears to lead to n–p cross sections showing considerably more asymmetry around 90° than is observed experimentally.

It may be possible by improved calculations, especially the odd triplet tensor scattering, to obtain better agreement. One attractive possibility of removing the first two of the above enumerated discrepancies is to include in the triplet central potential the repulsive core. Because of the limitation set by the effective range the radius of the repulsive core must be less than $0{\cdot}2 \times 10^{-13}$ cm. Even a repulsive core as small as this will lead to lower values for the n–p total cross section. Unfortunately, the magnitude of the angular variation will also be decreased. It will therefore be necessary to increase the long-tailed character of the tensor potential in order to account for the n–p angular variation at 28 and 40 Mev.

There is considerable evidence from other than high-energy scattering for the similarity between the n–n and p–p forces as required by the charge-independent hypothesis. This evidence is: (i) the binding energies and excited levels in mirror nuclei differ by an amount attributable to the difference in the electrostatic and magnetic interactions of the mirror nuclei (see especially Brown *et al.* 1951 for a brief review and additional references), (ii) the angular distribution of scattering and reaction products in mirror systems exhibit very similar behaviour (n–d and p–d scattering below 10 Mev (Breit 1950); the reaction products in the reactions $T(d, n)^4He$ and $^3He(d, p)^4He$ (Brolley, Fowler and Stovall 1951, Allred 1951)), and (iii) the γ-ray spectrum resulting from the reaction $\pi^- + d \rightarrow 2n + \gamma$ has been shown to depend strongly on the nature of the n–n interaction. The present experiments (Aamodt *et al.* 1951) indicate that an upper limit of 200 kev can be put on the binding energy of the di-neutron (Watson and Stuart 1951). This can be compared with the virtual n–p singlet level at 65 kev. It is difficult, however, to get a quantitative estimate from these data concerning the equality of the n–n and p–p interactions.

CONCLUSIONS

It is possible to get fair agreement with the experimental results using a static potential. The differences between theoretical predictions and experimental results amount to some 20%. This discrepancy, however, is much smaller than that resulting from the use of the static potentials suggested by meson theories. Probably better agreement can be obtained by small changes in the various parameters and more judicious choices of the radial forms. It is believed that there will be only slight changes in the overall picture, and a great deal has been learned concerning the potentials in the odd states.

It has not proved possible to find a charge-independent model that gives as good agreement as a charge-dependent model, but this cannot be interpreted as

showing that nuclear forces actually are charge dependent until further theoretical work is done. It would be desirable to ascertain with more certainty experimentally if n–p scattering is asymmetrical about 90°, for if there were no asymmetry this would strongly indicate that nuclear forces are charge-independent.

To some extent the hope that the high-energy experiments would lead to a comprehensive understanding of nuclear forces has not been fulfilled. This is due in part to the lack of a more general theory and the failure of the simple meson theories to give any kind of agreement. The lack of any clear mechanism to explain saturation by two-body forces alone makes a further uncertainty in the interpretation.

A fortiori we may explain the small role of large relativistic corrections by the fact that the discrepancies with experiments are just about as bad in the 28 MeV to 100 MeV region as in the 100 MeV to 340 MeV region. This does not indicate that relativistic corrections may not be important in a more complete theory but that the corrections may be 'mocked up' by a suitable potential interaction. The disagreement in the theoretical and experimental values of the total cross section may indicate the breakdown of the potential picture in S states, especially since large contributions of the S phase shift come from the interaction at very short distances ($\lesssim 0{\cdot}5 \times 10^{-13}$ cm).

Below we summarize state by state the nature of the potential operating in the state as has been determined by the present analysis:

n–p even parity triplet states: If the central and tensor forces have the same range, a long-tailed potential such as the Yukawa of range $1{\cdot}35 \times 10^{-13}$ cm with a central depth of 25·3 MeV and a tensor depth of 48·2 MeV, or the exponential potential of range $0{\cdot}75 \times 10^{-13}$ cm with a central depth of 69 MeV and a tensor depth of 128 MeV must be used. Of these two, the Yukawa potential gives the best fit to the angular distributions (particularly at the lower energies) and the exponential potential the best fit to the total cross section. The tensor range may be increased relative to the central range by a factor of two for the Yukawa potential. It might be desirable from the standpoint of the n–p total cross section to use a shallow type central interaction (or a repulsive core) and increase the tensor range to such an extent that the angular distribution predicted from the combination is in agreement with experiment.

n–p odd parity triplet states: It is definitely established that the amount of central scattering in the odd state must be a small fraction ($\lesssim 0{\cdot}1$) of the total scattering. Slightly better fits are obtained by employing a weak tensor potential in the odd states. The radial form best suited has not been determined from the present experiments. With the choice of the Yukawa potential having the same range as in the even states the depths must be less than 8–10 MeV. (The depth required for the best fit to the n–p data appears to be less than one half that required for the corresponding state in the p–p system.)

n–p even parity singlet states: The radial form is not well defined since the singlet scattering is fairly well masked by the triplet scattering. Best fits are, however, obtained with the long-tailed potentials. The best choice of the range and the depth for the potential is close to that required for the singlet p–p even states, although its direct determination from lower energy data leads to uncertainties of approximately $\pm 0{\cdot}5 \times 10^{-13}$ cm in the effective range.

n–p odd parity singlet states: There is no indication of any scattering in the odd states; however, a depth of 10% of that in the even states could probably be tolerated.

p–p even parity singlet states: The potential is cut off so that only a small D wave phase shift occurs at 32 MEV. The square well of range $2{\cdot}615 \times 10^{-13}$ cm and 13·273 MEV deep used with the repulsive core model in the triplet states is satisfactory in this respect.

p–p odd parity triple states: No indication of any scattering by either attractive or repulsive central forces. A tensor force is required that has a long tail and an extremely strong interaction at short distances if the square-well potential is used in the singlet states. If the repulsive core model is used in the singlet state the triplet tensor potential need not have such a strong interaction at short distances, and an exponential radial form is acceptable. The sign of the tensor scattering is not determined from the p–p interaction alone.

The explanation of p–p scattering by the use of either the repulsive core model or the singular tensor model requires a strong interaction at short distances.

ACKNOWLEDGMENTS

I wish to thank Professor Serber for valuable discussions preliminary to the preparation of this review, and to acknowledge again his basic contributions to the research papers in which I had a part. I am also grateful to Donald Swanson for suggestions and for calculating some of the angular distributions used in the present report.

REFERENCES

AAMODT, R. L., PANOFSKY, W. K. H., and PHILLIPS, R., 1951, *Phys. Rev.*, **83,** 1057.
ALLRED, J. C., 1951, *Phys. Rev.*, **84,** 695.
ASHKIN, J., and WU, T., 1948, *Phys. Rev.*, **73,** 973.
BARKER, F. C., and RAVENHALL, D. G., 1949, *Nature, Lond.*, **163,** 20.
BELL, R. E., and ELLIOTT, L. G., 1950, *Phys. Rev.*, **79,** 282.
BETHE, H. A., 1949, *Phys. Rev.*, **76,** 38.
BETHE, H. A., and LONGMIRE, C., 1950, *Phys. Rev.*, **77,** 647.
BIRGE, R. W., KRUSE, U. E., and RAMSEY, N. F., 1951, *Phys. Rev.*, **83,** 274.
BLATT, J. M., 1948, *Phys. Rev.*, **74,** 92.
BRADNER, H., CRAWFORD, R., GORDON, H., and WOODYARD, J. R., 1948, *Phys. Rev.*, **73,** 534.
BREIT, G., 1938, *Phys. Rev.*, **53,** 153; 1950, *Ibid.*, **80,** 1110.
BREIT, G., and BLOCH, I., 1947, *Phys. Rev.*, **72,** 135.
BREIT, G., CONDON, E. U., and PRESENT, R. P., 1936, *Phys. Rev.*, **50,** 825.
BREIT, G., HOISINGTON, L. E., and SHARE, S. S., 1939, *Phys. Rev.*, **56,** 884.
BREIT, G., THAXTON, H. M., and EISENBUD, L., 1939, *Phys. Rev.*, **55,** 1018.
BROLLEY, J. E., Jr., COON, J. H., and FOWLER, J. L., 1951, *Phys. Rev.*, **82,** 190.
BROLLEY, J. E., Jr., FOWLER, J. L., and STOVALL, E. J., Jr., 1951, *Phys. Rev.*, **82,** 502.
BROWN, A. B., SNYDER, C. W., FOWLER, W. A., and LAURITSEN, C. C., 1951, *Phys. Rev.*, **82,** 159.
BRUECKNER, K., HARTSOUGH, W., HAYWARD, E., and POWELL, W. M., 1949, *Phys. Rev.*, **75,** 555.
BURHOP, E. H. S., and YADAV, H. N., 1949, *Proc. Roy. Soc.* A, **197,** 505.
CASE, K. M., and PAIS, A., 1950, *Phys. Rev.*, **80,** 203.
CASTILLEJO, L., and RICHARDSON, H. T , 1949, *Phys. Rev.*, **76,** 1732.
CHAMBERLAIN, O., SEGRÈ, E., and WIEGAND, C., 1951 a, *Phys. Rev.*, **81,** 284; 1951 b, *Ibid.*, **81,** 661; 1951 c, *Ibid.*, **83,** 923
CHAMBERLAIN, O., and WIEGAND, C., 1950, *Phys. Rev.*, **79,** 81.
CHEW, G. F., and GOLDBERGER, M. L., 1949, *Phys. Rev.*, **75,** 1637.
CHRISTIAN, R. S., 1949, *Phys. Rev.*, **75,** 1675.
CHRISTIAN, R. S., and HART, E. W., 1950, *Phys. Rev.*, **77,** 441.
CHRISTIAN, R. S., and NOYES, H. P., 1950, *Phys. Rev.*, **79,** 85.
CHUPP, W. W., GARDNER, E., and TAYLOR, T. B., 1948, *Phys. Rev.*, **73,** 742.
COOK, L. J., MCMILLAN, E. M., PETERSON, J. M., and SEWELL, D. C., 1947, *Phys. Rev.*, **72,** 1264.

Cork, B., 1950, *Phys. Rev.*, **80**, 321.
Cork, B., Johnston, L., and Richman, C., 1950, *Phys. Rev.*, **79**, 71.
DeJuren, J., 1950, *Phys. Rev.*, **80**, 27.
DeJuren, J., and Knable, N. 1950, *Phys. Rev.*, **77**, 606.
DeJuren, J., and Moyer, B. J., 1951, *Phys. Rev.*, **81**, 919.
Demeur, M., 1950, *Physica*, **16**, 353.
Fox, R., Leith, C., and Wouters, L., 1950, *Phys. Rev.*, **80**, 23.
Gerjouy, E., 1950, *Phys. Rev.*, **77**, 568.
Hadley, J., Kelley, E., Segrè, E., Wiegand, C., and York, H., 1949, *Phys. Rev.*, **75**, 351.
Heckrotte, W., and Wolff, P., 1948, *Phys. Rev.*, **73**, 265.
Hefner, W., and Peierls, R., 1942, *Proc. Roy. Soc.* A, **181**, 43.
Helmholz, A. C., McMillan, E. M., and Sewell, D., 1947, *Phys. Rev.*, **72**, 1003.
Hildebrand, R. H., and Leith, C. E., 1950, *Phys. Rev.*, **80**, 842.
Hsu, K. N., and Hu, T. M., 1949, *Phys. Rev.*, **75**, 987.
Hu, T. M., and Massey, H S. W., 1949, *Proc. Roy. Soc.* A, **196**, 135.
Jackson, J. D., and Blatt, J. M., 1950, *Rev. Mod. Phys.*, **22**, 77.
Jastrow, R., 1951, *Phys. Rev.*, **81**, 165.
Jost, R., and Pais, A., 1951, *Phys. Rev.*, **82**, 840.
Kelly, E., Leith, C., Segrè, E., and Wiegand, C., 1950, *Phys. Rev.*, **79**, 96.
Lampi, E. E., Freier, G. D., and Williams, J. H., 1950, *Phys. Rev.*, **80**, 853.
Leith, C. E., 1950, *Phys. Rev.*, **78**, 89.
Lippmann, B. A., and Schwinger, J., 1950, *Phys. Rev.*, **79**, 469.
Melkonian, E., 1949, *Phys. Rev.*, **76**, 1744.
Mobley, R. C., and Laubenstein, R. A., 1950, *Phys. Rev.*, **80**, 309.
Nordsieck, A., 1940, *Phys. Rev.*, **58**, 310.
Oxley, C. L., Schamberger, R. D., and Towler, O. A., Jr., 1951, *Phys. Rev.*, **82**, 295.
Panofsky, W. K. H., and Fillmore, F. L., 1950, *Phys. Rev.*, **79**, 57.
Pease, R. L., and Feshbach, H., 1950, *Phys. Rev.*, **78**, 322.
Primakoff, H., 1947, *Phys. Rev.*, **72**, 118.
Ringo, G. R., and Burgy, M. T., 1951, *Phys. Rev.*, **82**, 344.
Roberts, T. R., and Nier, A. O., 1950, *Phys. Rev.*, **77**, 746.
Rohrlich, F., and Eisenstein, J., 1949, *Phys. Rev.*, **75**, 705.
Rosenfeld, L., 1948, *Nuclear Forces* (Amsterdam : North Holland Publishing Co.; New York : Interscience).
Sachs, R. G., 1947, *Phys. Rev.*, **72**, 91.
Salpeter, E. E., 1951, *Phys. Rev.*, **82**, 60.
Schwinger, J., 1947 a, Hectographed notes on *Nuclear Physics*, Harvard; 1947 b, *Phys. Rev.*, **72**, 742; 1950, *Ibid.*, **78**, 135.
Serber, R., 1947, *Phys. Rev.*, **72**, 1008.
Sherr, R., 1945, *Phys. Rev.*, **68**, 240.
Siegal, A., 1951, *Phys. Rev.*, **82**, 194.
Sleator, W., Jr., 1947, *Phys. Rev.*, **72**, 207.
Snyder, H., and Marshak, R. E., 1947, *Phys. Rev.*, **72**, 1253.
Swanson, D. R., 1951, *Phys. Rev.*, **84**, 1068.
Taschek, R. F., Argo, H. V., Hemmendinger, A., and Jarvis, G. A., 1949, *Phys. Rev.*, **76**, 325.
Taylor, A. E., Pickavance, T. G., Cassels, J. M., and Randle, T. C., 1951, *Phil. Mag.*, **42**, 20.
Wallace, R., 1951, *Phys. Rev.*, **81**, 493.
Watson, K. M., and Lepore, J. V., 1949, *Phys. Rev.*, **76**, 1157.
Watson, K. M., and Stuart, R. W., 1951, *Phys. Rev.*, **82**, 738.
Wiegand, C., 1950, *Rev. Sci. Instrum.*, **21**, 975.
Wolfenstein, L., 1949, *Phys. Rev.*, **76**, 541; 1951, *Ibid.*, **82**, 308.
Wouters, L., 1951, *Phys. Rev.*, **84**, 1069.
Yadav, H. N., 1949, *Phys. Rev.*, **76**, 1720.

TURBULENT MOTION

By G. K. BATCHELOR
Trinity College, Cambridge

CONTENTS

	PAGE
§ 1. Introduction	101
§ 2. The statement of the problem	102
§ 3. General ideas about turbulent motion	105
§ 4. The universal equilibrium of the small-scale components	108
(i) The theory	108
(ii) Experimental evidence	111
(iii) Applications of the theory	114
§ 5. The effect of turbulent motion on the distribution of local properties of the fluid	115
§ 6. Homogeneous decaying turbulence	118
§ 7. Flows in which there is a mean shearing motion	122
§ 8. Miscellaneous developments	128
(i) The passage of homogeneous turbulence through gauze	128
(ii) Hypotheses about the transfer of energy across the spectrum	128
(iii) The probability distribution of the velocity fluctuations	130
(iv) Generation of noise by turbulent motion	133
Appendix (by Dr. A. A. Townsend): Advances in measuring technique	135
References	140

ABSTRACT. This report describes briefly the recent advances in the study of turbulent motion that seem to the author to be significant. Most of these advances have occurred since about 1935 and are still being pursued actively, since turbulent motion is not yet understood properly. The emphasis is on the physical content of the various ideas and results rather than on their mathematical formulation.

A prominent place is given to the universal similarity theory since this seems likely to provide a valid (although partial) description of the small-scale components of the motion. A section is devoted to a problem which occurs in many different physical fields, viz. the effect of turbulent convection on the distribution of various local properties of the fluid. The recent work on the study of homogeneous decaying turbulence is summarized. The very important and extensive problem of turbulent shear-flow is considered, although it has not been possible to do justice to all the available experimental results. There is an Appendix on the advances in measuring technique that have made possible much of the recent research.

§ 1. INTRODUCTION

THE subject of turbulent motion of fluids has appeared in two previous *Reports on Progress in Physics*. In the very first volume Trubridge (1934) gave an account of an attempt by Burgers to apply the methods of statistical mechanics to the problem, and two years later Fage (1936) described the current ideas about turbulence. The literature on turbulent motion has been more than doubled since that time, and there has been a corresponding amount of progress. The air of mystery that surrounds the phenomenon of turbulence has not yet been entirely dispelled, but it has been considerably reduced. We can state what we do not know much more closely than we could sixteen years ago.

In this article the emphasis will be on the general physical ideas that have been put forward, rather than on their mathematical consequences. This is no concession to any lack of mathematical interest on the part of readers; the mathematics *is* less important than the physics, since the problem of turbulence is at the primitive stage in which differences of opinion concern assumptions rather than conclusions. It is commonly said that the test of a theory is that it should lead to predictions which agree with measurements, but this is only the final test. A more elementary and more important test is that it should be rational, self-consistent, plausible, and it is principally on the basis of this test that it is appropriate to judge current theories of turbulence. Later, when there is more general agreement about the broad features of the problem, we shall need to make greater use of the more discriminating test of agreement between detailed predictions and experiment. Meanwhile we can eschew elaborate mathematics unless we like it for its own sake.

The essentially physical nature of the problem of turbulence has been widely appreciated in recent years, and a distinguished list of physicists have made contributions to the subject. Simultaneously (and doubtless not independently) the problems of physics in which turbulent motion is recognized as playing a part have been multiplying. Heat exchange in the atmosphere, the scintillation of star-light, diffusion of radioactive substances, ocean currents, the scouring of river beds, the motion of interstellar gas clouds (and plasma oscillations?)—in all these problems a knowledge of the characteristics of turbulent motion is becoming more and more desirable. It is of course not surprising that turbulence is so ubiquitous. Given that matter is in either a liquid or gaseous form (or, more precisely, given that it can be regarded as a fluid for dynamical purposes), and given that the Reynolds number is large (as it always is, with air or water, for motion on anything larger than a laboratory scale), it is virtually inevitable that any kinetic energy of the matter will exist as a turbulent motion. Any other form of motion would almost certainly be unstable.

This interest in the properties of turbulent motion—part intrinsic, part utilitarian—on the part of physicists has left a definite mark on the developments of the last few years. At the time when turbulence was of importance chiefly from the point of view of aeronautics, interest was concentrated on such problems as the skin friction on a surface due to a turbulent boundary layer, or the way in which turbulence in a stream affects laminar boundary layers. The more recent trend is to take a more fundamental viewpoint and to regard turbulent motion in general as a statistical mechanical system. The approach has been well justified, since it now seems that there are certain statistical features which are common to all kinds of turbulent motion. The practical questions put by aeronautical engineers and meteorologists and others, for example about the magnitude of the skin friction at a boundary, are still without an answer, but the clarity brought to the subject in recent years gives hope that answers will soon be obtained.

§2. THE STATEMENT OF THE PROBLEM

We must be clear about what we have in mind when we speak of the 'problem of turbulent motion'. In general we are referring to a uniform Newtonian fluid, the motion of which satisfies the conditions of conservation of mass and momentum. In the simple and common case in which the relative velocities in the fluid are small compared with the velocity of sound, we can regard the fluid

as effectively incompressible, and the analytical expression of the above conditions is then as follows:

$$\nabla.\mathbf{v}=0, \qquad \ldots\ldots(2.1)$$

$$\frac{\partial \mathbf{v}}{\partial t}+\mathbf{v}.\nabla\mathbf{v}=-\frac{1}{\rho}\nabla p+\nu\nabla^2\mathbf{v}, \qquad \ldots\ldots(2.2)$$

where $\mathbf{v}$ and p are the (vector) velocity and hydrostatic pressure at any point of the fluid, ρ is the density and $\nu(=\mu/\rho)$ is the kinematic viscosity.

The above pair of equations possess many different solutions, each of which describes a field of flow; the sciences of hydrodynamics, aerodynamics, and hydraulics are largely based on certain categories of these solutions. Our concern is with a particular class of solutions, viz. those that exhibit a random variation of the velocity with respect to time t and the (vector) position variable $\mathbf{x}$. 'Random' is the keyword here, and it supplies a necessary and sufficient description of turbulent motion. Nevertheless it is not easy to give a precise definition of a random variation of the velocity. It is necessary to adopt the procedure of imagining a statistical ensemble of solutions, all of which satisfy the given boundary and initial conditions. This statistical ensemble defines the probability that $\mathbf{v}(\mathbf{x}, t)$ will lie within specified limits, and likewise it defines the statistical connection between the velocities of the fluid at different positions and times. In other words, we are supposing, as a matter of definition of turbulent motion, that the data of the problem (i.e. the governing equations and the given boundary and initial conditions) are not sufficient to determine $\mathbf{v}$ as a function of $\mathbf{x}$ and t, but are sufficient to determine the probability laws which describe the values of $\mathbf{v}$. Thus if the experiment of producing the flow field with the required initial and boundary conditions were performed many times, each experiment supplying one 'realization' of the flow field, it would be found that the value of $\mathbf{v}$ at given values of $\mathbf{x}$ and t would vary from one experiment to the next, and the aggregate of the values of $\mathbf{v}$ would define (asymptotically) the probability distribution of $\mathbf{v}(\mathbf{x}, t)$.

The velocity field is of course wholly determined when sufficient information is supplied; the above insufficiency of the data has its origin in the ignoring of small disturbances in the initial and boundary conditions which are outside experimental control and which vary from one experiment to the next. To clarify the matter, let us consider a particular flow problem. Suppose the space between two concentric spheres is filled with fluid and the spheres are set into steady relative angular motion. In the idealized case in which a true fluid, initially at rest, moves in the space between perfectly smooth rigid walls which are in absolutely steady motion, there is no arbitrariness about the velocity of the fluid, and turbulent motion cannot occur. However, this laminar motion is unstable (when the Reynolds number is sufficiently large), and if, for instance, the outer sphere is not quite rigid and it transmits small irregular disturbances—arising from any source extraneous to the flow system—the laminar motion will break down and a turbulent motion will be set up. If we could specify these small extraneous disturbances and use them as part of the boundary conditions for the motion of the fluid, the consequent turbulent motion could be computed (in principle). However, the extraneous disturbances cannot be controlled at will, so we call them 'random', and we give up the intention of computing the turbulent motion in detail. Instead, we make use of the experimental observation that *the probability laws describing the turbulent motion are independent of the small*

extraneous disturbances from which the turbulence arises and depend only on the given data of the problem. This remarkable fact shows that we can ignore the extraneous disturbances and concentrate on finding the probability laws of the turbulent motion as a unique function of the initial and boundary conditions of the problem.

A parallel to the above situation can be found in the kinetic theory of gases. The medium is not here continuous, but consists of a large number of discrete particles whose collisions satisfy certain equations which can be regarded as analogues of (2.1) and (2.2). When the initial positions and motions of the particles are given and the conditions at the boundary of the gas are known, the subsequent history of the particles can be computed. In certain freak cases the motion is ordered and persistent; for instance, if all the particles initially move in a direction perpendicular to two perfectly reflecting walls they will continue to do so (the analogous laminar motion of a fluid is by no means so rare, owing to the damping action of viscosity). However, we know that in practice such ordered motions will always be disturbed, and we know also (from both experiment and theory) that the disturbance produces a 'random' motion described by probability laws (such as a Maxwellian distribution of the velocities) which are independent of the nature of the disturbance. Provided the motion of the particles is such that they are colliding, we know that the statistical laws are applicable. The same is true of the motion of the fluid. Provided the various degrees of freedom of the motion (a concept which will be discussed later) are interacting, the motion can be called turbulent and can be described by statistical relations.

Hence the formal statement of the problem to be considered is this: Given a fluid in which motions conform to the equations (2.1) and (2.2), and given that the fluid is set into a turbulent motion which is determined statistically by the initial and boundary conditions and the equations (2.1) and (2.2), to find the probability laws describing the turbulent motion and their variation with time.

The value of making such a formulation of the problem does not lie in the practical guidance which it supplies—it is too formal for that—but in the clarification of the theoretical basis of the problem. It is evident, for instance, that the various mean values used to describe the turbulence should be defined as integrals with respect to the appropriate probability distribution. If averages over large intervals of time or space are obtained experimentally, it will be possible to identify these with the theoretical averages only if an appropriate ergodic theorem can be established. It is a known result* that probability and physical averages are identical when the random function, viz. the velocity $\mathbf{v}(\mathbf{x}, t)$, is statistically stationary with respect to the physical variable concerned (and satisfies certain other, weak, conditions). Hence, if the turbulence is statistically steady, an average over a long interval of time is identical with the theoretical probability average, and if the turbulence is statistically uniform in certain spatial directions the same is true of an integral over a long interval of the appropriate space coordinate. In the case of turbulent flow in a pipe under a steady pressure gradient, experimental averages over either time or distance in the direction of the pipe can replace the theoretical probability average.

Nothing has yet been said about the form of the 'probability laws' which are to be used to describe the turbulent motion. We can refer to the mathematical theory of random functions for the statement that a random function $\mathbf{v}(\mathbf{x}, t)$ is

* For the precise mathematical statement see Khintchine (1933).

fully determined statistically by the set of joint probability distributions of the values of $\mathbf{v}$ at any n pairs of values of $\mathbf{x}$ and t—in other words, by the set of all the mean values of products which can be formed from values of the velocity at different positions and times. A complete statistical specification of a turbulent motion thus requires a vast array of information. Any theory of turbulent motion which fails to take all these mean values into account, directly or indirectly, is bound to be faced with an insufficiency of dynamical equations, as has occasionally been noticed. Hitherto theories of turbulence have not attempted to be complete, so that the problem of handling the complete statistical specification of the motion has not been given much consideration. The working tools of the subject are the simpler velocity product mean values; these will be introduced, as needed, in the following sections.

§3. GENERAL IDEAS ABOUT TURBULENT MOTION

In the welter of contributions to different aspects of the subject there can be discerned a few general ideas that are finding common acceptance. One of the most important of these is the notion that turbulent motion consists of the superposition of a large number of component motions which differ from each other in their length scale. The equation of motion (2.2) is non-linear, so that the various component motions interact, and the observed properties of the turbulence are thought of as being the statistical result of such interaction. This conception of the motion does not, by itself, lead immediately to any results, but it has the considerable advantage of permitting the use of intuitive ideas about the behaviour of dynamical systems with a large number of degrees of freedom.

The resolution of the turbulence into a large number of component motions of different scale can be put on an exact basis and is not in any sense an assumption; what are less straightforward, as we shall see later, are the physical arguments erected on the basis of such a resolution. The resolution is in fact simply a resolution of the instantaneous distribution of the velocity distribution into orthogonal modes consistent with the boundary conditions. The instantaneous distribution of velocity along an elastic string stretched between two points distance l apart can be regarded as the sum of a series of Fourier components which make additive contributions to the kinetic energy, the fundamental wavelength being $2l$; likewise the instantaneous distribution of fluid velocity over a three-dimensional field can be regarded as the sum of a triple series of Fourier components. If the field extends to infinity the corresponding series becomes an integral, and if the boundaries have curvilinear form some other set of orthogonal functions (e.g. Bessel functions) may be appropriate, but these are modifications which do not affect the principle of the resolution. In the case in which the field extends to infinity and the velocity does not tend to zero (or a constant value) there, the mathematical refinements* to the usual Fourier resolution are not simple and will not be considered here. For our purposes it is sufficient to remark that a resolution of the general form

$$\mathbf{v}(\mathbf{x}) = \int \mathbf{A}(\boldsymbol{\kappa}) e^{i\boldsymbol{\kappa}\cdot\mathbf{x}} d\boldsymbol{\kappa}, \qquad (i = \sqrt{-1}) \qquad \ldots\ldots (3.1)$$

where the integration is over all $\boldsymbol{\kappa}$-space, can always be made, and that although the (vector) Fourier coefficient $\mathbf{A}(\boldsymbol{\kappa})$ diverges in certain cases for which either a

* There is an extensive literature on generalized harmonic analysis; two papers of broad scope are Wiener (1930) and Moyal (1949).

series or a Stieltjes integral would be more appropriate, we can continue to use the function heuristically without grave risk of falling into error (as in fact has been done by many authors—for an example of the use of the more rigorous mathematical technique, see Obukhoff (1941)).

$\mathbf{A}(\mathbf{x})e^{i\mathbf{x}\cdot\mathbf{x}}d\mathbf{x}$ is now the contribution to the velocity from the volume element $d\mathbf{x}$ of wave-number space and is the typical component of the turbulence, and represents a sinusoidal velocity distribution with wavelength $2\pi/\kappa$ ($\kappa = |\mathbf{x}|$), which is thus a measure of the length-scale of the component motion. The continuity equation (2.1) requires

$$\mathbf{x}\cdot\mathbf{A}(\mathbf{x}) = 0, \qquad \ldots\ldots(3.2)$$

so that the component motions are sinusoidal 'waves' with velocity parallel to $\mathbf{A}(\mathbf{x})$ and spatial variation in the direction of the wave numbers $\mathbf{x}$. (Note that these 'waves', unlike those on a stretched string, do not propagate, in view of the different form of the equation of momentum.) The square of the Fourier coefficient $\mathbf{A}(\mathbf{x})$ is proportional to the amount of kinetic energy associated with the Fourier component of wave number $\mathbf{x}$, which is also the contribution that the component makes to the total kinetic energy of the turbulence. The Fourier resolution of the instantaneous velocity field thus supplies a spectral distribution of kinetic energy; since the quantity $\mathbf{A}(\mathbf{x})$ fluctuates—i.e. it is different for different realizations of the velocity field—we take the average and finally obtain $\overline{\mathbf{A}^2(\mathbf{x})}$ as a measure of the distribution in wave-number space of the density of contributions to the kinetic energy of the turbulence.

So far as the above remarks have gone, the notion of a Fourier resolution of the velocity field does not differ in essential respects from that of an analysis of white light, or random sound, or electronic 'noise'. The essential differences enter when we consider the governing equation (2.2) which describes the relation between the spatial distributions of velocity at different instants of time. The rate of change of the Fourier coefficient $\mathbf{A}(\mathbf{x})$ can be obtained by substituting (3.1) in (2.2) and eliminating the pressure term with the aid of (2.1):

$$\frac{\partial\mathbf{A}(\mathbf{x})}{\partial t} = i\int \mathbf{x}'\cdot\mathbf{A}^*(\mathbf{x}'-\mathbf{x})\left[\mathbf{x}\cdot\mathbf{A}(\mathbf{x}')\frac{\mathbf{x}}{\kappa^2} - \mathbf{A}(\mathbf{x}')\right]d\mathbf{x}' - \nu\kappa^2\mathbf{A}(\mathbf{x}), \quad \ldots\ldots(3.3)$$

where * denotes a complex conjugate. The term arising from the effect of viscosity represents a linear damping action, as we should expect, and shows that the dissipation of energy is more rapid (relatively) for the small-scale than for the large-scale components. The term arising from the inertia and pressure forces, on the other hand, is non-linear and shows that there is a continual interaction, or modulation, between the different Fourier components. The variation of the energy of the typical Fourier component is given by

$$\frac{\partial\mathbf{A}(\mathbf{x})\cdot\mathbf{A}^*(\mathbf{x})}{\partial t} = i\int[-\mathbf{x}'\cdot\mathbf{A}^*(\mathbf{x}'-\mathbf{x})\mathbf{A}^*(\mathbf{x})\cdot\mathbf{A}(\mathbf{x}') + \mathbf{x}'\cdot\mathbf{A}(\mathbf{x}'-\mathbf{x})\mathbf{A}(\mathbf{x})\cdot\mathbf{A}^*(\mathbf{x}')]\,d\mathbf{x}' - 2\nu\kappa^2\mathbf{A}(\mathbf{x})\cdot\mathbf{A}^*(\mathbf{x}), \quad \ldots\ldots(3.4)$$

showing that the interaction leads, in general, to a transfer of energy between any two components. It is this inertial interaction between the components, in any one of its many forms, that is the principal obstacle to a solution of the problem of turbulent motion. A solution of equation (3.3) does not seem to be within our reach, any more than is a solution of (2.2) (except in certain simple cases of laminar motion), and 'theories' of turbulent motion consist essentially of hypotheses which permit a quantitative consideration of the interaction between the components for whatever limited purpose may be under discussion.

If we think of the turbulent motion as a statistical dynamical system with the Fourier components as degrees of freedom, equation (3.3) is the counterpart of a 'collision' law. Excitation of any one degree of freedom leads, in general, to excitation of all other degrees of freedom to an extent which is limited only by the damping action of viscosity. In view of the existence of the dissipation, there is no question here of using the principle of equipartition of energy in any form. The equations of motion are not Hamiltonian, and there is no counterpart of the Liouville theorem of classical statistical mechanics, so that even in the case of vanishing viscosity we do not expect to find equipartition of energy among the various degrees of freedom.

The single parameter which enters into equations (2.2) and (3.3) is the viscosity ν, or alternatively, if we make all quantities dimensionless with the aid of a reference velocity V and a length L, it is the Reynolds number VL/ν. We thus expect to find a different turbulent motion at each different Reynolds number. In view of the fact that ν enters into equation (3.4) in association with the factor κ^2, the effect of viscous forces will always be dominant for Fourier components with sufficiently large wave numbers, but the region of strong influence of viscosity moves outward from the origin of wave-number space as the Reynolds number increases (i.e. as ν decreases). At lower wave numbers the rate of change of the Fourier coefficients is dominated by inertia effects; this part of the dynamical system is thus non-dissipative but it is not closed, and in general there will be a statistical flow of energy (i.e. a biased exchange) towards that part of wave-number space where energy is being absorbed by viscosity. It is profitable to think about the simple limiting case of infinite Reynolds number or vanishing viscosity. The region of wave-number space in which viscous damping is significant is now displaced to $\kappa = \infty$, and a 'violet catastrophe' lies in wait for any energy which is transferred, by inertia effects, out of the region in which κ is finite. This simple case in which only inertia forces operate, except at $\kappa = \infty$, offers the best opportunity for the use of intuitive hypotheses about the interaction between components. The expectation we have is that the energy spectrum of the turbulence has a shape at small values of κ which is influenced strongly by the boundary conditions, and that at larger values of κ it falls off with increasing κ in a way which is governed by the inertial exchange of energy. At infinite Reynolds number this falling-off continues to $\kappa = \infty$, but at finite Reynolds number a value of κ will be reached at which the viscous damping effect is appreciable, and the spectrum will fall off to zero much more rapidly in the neighbourhood of this wave number.

Related information about the velocity distribution at very large Reynolds numbers can be obtained from a consideration of the form of the differential equation (2.2), as has been made clear by Burgers (1948). The highest order differential in this equation is multiplied by the small parameter ν. Now Burgers has shown from an examination of similar (but simpler) differential equations that the effect of the convective term $\mathbf{v}.\nabla\mathbf{v}$ is to produce increasingly sharp gradients of the velocity at various points in the fluid; both shock waves and boundary layers may be regarded as products of this process. An opposing tendency to diminish the gradients is supplied by the term $\nu\nabla^2\mathbf{v}$, but if ν is small the smoothing effect becomes significant only when the derivative $\nabla^2\mathbf{v}$ has been built up to a large value. Hence we expect that when a fluid is in turbulent motion at large Reynolds numbers there will exist in the fluid a number of thin

'dissipation layers' inside which the velocity varies rapidly (with position), and in which the opposing tendencies of the terms $\mathbf{v}.\nabla^2\mathbf{v}$ and $\nu\nabla^2\mathbf{v}$ are in approximate balance. When we recall that near-discontinuities in the spatial distribution of $\mathbf{v}$ have a strong influence on the higher harmonics of a Fourier analysis of $\mathbf{v}$, the relation between the remarks in this and the preceding paragraph becomes evident.

§4. THE UNIVERSAL EQUILIBRIUM OF THE SMALL-SCALE COMPONENTS

(i) *The Theory*

The notions explained in the preceding section have been developed chiefly in association with the very important theory of universal equilibrium (sometimes given the alternative description of universal similarity) of the Fourie components of large wave number. This theory was put forward first by Kolmogoroff (1941 a, b) and Obukhoff (1941), and different versions were later suggested, independently, by Onsager (1945) and von Weizsäcker (1948). Since the first critical review (Batchelor 1947) there have been many attempts to assess the validity of the theory, and although there is still doubt about the quantitative conditions which must first be satisfied it is now widely held to be substantially correct. Only the essential ideas of the theory will be presented here.

It has been remarked that the effect of inertia forces is to excite, or transfer energy to, as many Fourier components as is consistent with the damping effect of viscosity. Let us suppose that the length-scale representative of the boundaries is L. In general this will be the length-scale of Fourier components of the motion that are directly excited by the action of the boundaries; for example, when fluid flows through a pipe of diameter L under pressure, the Fourier components of the velocity distribution that are directly influenced by the boundary conditions will have wave numbers which are of the order of magnitude of $1/L$. Once these components of the turbulent motion are excited, inertia forces operate to spread the energy over a wider range of the spectrum and thus, in general, to transfer it to larger wave numbers. As already explained, this tendency will be checked only when the energy reaches wave numbers which are so large that viscous damping is appreciable.

It is clear that, if we plan to isolate certain parts of a dynamical system with the intention of gaining a limited understanding of the motion thereby, it will be useful to think about the Fourier components with wave numbers which are large compared with $1/L$, since these components are excited by the action of inertia forces alone and are not subject to the direct action of the boundaries. If, as seems probable, the influence of the boundary conditions is gradually lost in the statistical process of transfer of energy to larger wave numbers—corresponding to the intuitive idea that a statistical transfer must be accompanied by a loss of information, and called the 'disorder' hypothesis by von Weizsäcker—the motion due to Fourier components of sufficiently large wave number will be statistically independent of the boundary conditions and of the source of the kinetic energy. This region of the spectrum should therefore present a relatively simple problem, which is not complicated by the specification of the exact shape and motion of the boundaries. These Fourier components with wave numbers κ such that $\kappa \gg 1/L$ (where the numerical interpretation of $\gg$ depends on how rapidly the influence of the boundary conditions is eliminated in the process of transfer of energy across the spectrum) are the subject of the universal equilibrium theory.

Evidently it is necessary that the wave number at which viscous damping produces a rapid decrease in the spectral distribution of energy ($=1/\eta$ say, where η has the meaning of a length characteristic of the smallest eddies, or component motions, existing in the turbulence) must also satisfy this condition, for otherwise the Fourier components to which the disorder hypothesis applies would not be excited. Since, as we saw in the preceding section, increase of the Reynolds number VL/ν has the effect of decreasing η, we must restrict our discussion to cases of turbulent motion for which VL/ν is so large that

$$\eta \ll L, \qquad \ldots\ldots (4.1)$$

that is, that the smallest eddies of the turbulence are small compared with the largest eddies.

In general, turbulent motions are such that the statistical quantities describing the motion (mean-square velocity, etc.) vary with position in the fluid. The scale of this non-uniformity is determined by the geometry of the boundaries, and in general is of the order of magnitude of L. Consequently the variation of the statistical quantities over a region with linear dimensions small compared with L is negligible, and we can regard the turbulence as spatially homogeneous within this region. If the distribution of velocity fluctuation (about the spatial average) over this region is analysed, we shall obtain a set of Fourier coefficients which have *local* significance and all of which refer to wave numbers large compared with $1/L$. All that has been said hitherto about Fourier coefficients of large wave numbers applies either to a local or to the whole velocity field, but it is more convenient to confine attention now to the Fourier coefficients, of large wave numbers, of the distribution of velocity in the neighbourhood of any point in the fluid.

The turbulence can likewise be regarded as statistically uniform over suitably small intervals of time. With the weak assumption that the small-scale Fourier components have relatively small characteristic times (that is, that small eddies have small periods, or large vorticity), we see that the small-scale Fourier components will be unaffected by slow temporal variations of the mean quantities describing the turbulence and will depend only on their instantaneous values. Speaking dynamically, we can say that the degrees of freedom corresponding to large values of κ are always in statistical equilibrium on account of their rapid response to external conditions. Energy is transferred to them from components with smaller wave numbers, but it is lost by transfer to components with larger wave numbers (and perhaps by dissipation) at an equal rate.

Thus we have reached the useful conclusion that when the Reynolds number is such that the condition (4.1) is satisfied, the Fourier coefficients (of the velocity distribution in the neighbourhood of any point) belonging to wave numbers $\kappa \gg 1/L$ are statistically independent of the boundary conditions and are determined by the local (in space and time) mean properties of the turbulence. In particular we may notice that the directional properties imposed on the turbulence as a whole by the boundary conditions cannot be transmitted to the small-scale components, according to the 'disorder' hypothesis, so that the motion associated with these small-scale components must be statistically isotropic.

The argument now turns in the reverse direction and establishes the quantities on which the small-scale Fourier components *do* depend. These small-scale components—belonging to what we shall call the equilibrium range of wave

numbers—owe their excitation to the inertial transfer of energy from the large-scale components of the turbulence, and the intensity of this excitation will clearly depend on the rate at which this transfer of energy proceeds. In view of the fact that *all* the viscous dissipation occurs in the equilibrium range (this is the meaning of the condition (4.1) on the Reynolds number), the rate of transfer of energy from the large-scale components to the components in the equilibrium range must be equal to the rate of dissipation of energy ($=\epsilon$, say). This quantity ϵ (which must be local, in space and time) seems to be the only one needed to express the influence of the field of motion as a whole on the equilibrium range of Fourier components near any point in the fluid. Once the local value of ϵ has been specified, the small-scale components adjust themselves so as to receive energy at this rate and so as to dissipate it. The details of the dissipation process will depend on the value of ν, of course. Apart from the dependence of these two parameters ϵ and ν, the motion associated with the equilibrium range of wave numbers (i.e. the relative motion in the neighbourhood of any point) will be statistically universal. This is the principal prediction of the theory.

In the case in which the Reynolds number is larger, by an order of magnitude, than that needed for the above remarks to apply, the theory can be taken further. The lowest wave number in the equilibrium range is presumably determined by the details of the disordering process which accompanies transfer of energy across the spectrum, and is independent of Reynolds number. Increase of the Reynolds number causes an increase in the highest wave number in the equilibrium range (viz. $1/\eta$), and hence a corresponding increase in the *extent* of the equilibrium range. When the Reynolds number is so large that the Fourier coefficients at one end of the equilibrium range are statistically independent of those at the other end, the condition for which will be that it is possible to find wave numbers κ such that

$$1/L \ll \kappa \ll 1/\eta, \qquad \ldots\ldots(4.2)$$

the Fourier coefficients at the smaller wave numbers within the equilibrium range will no longer be influenced by viscous dissipation and will not depend on the parameter ν. The statistical flow of energy into any part of this inertial subrange from smaller wave numbers is balanced by an outflow to higher wave numbers at an equal rate (viz. ϵ), and the motion associated with the subrange is uniquely determined statistically by the single parameter ϵ.

It is now a simple matter to work out the predictions of the theory for particular statistical quantities. Any mean value which is determined by the Fourier components of the equilibrium range must have an isotropic form which is determined uniquely by the parameters ϵ (dimensions $[\mathrm{L}]^2[\mathrm{T}]^{-3}$) and ν (dimensions $[\mathrm{L}]^2[\mathrm{T}]^{-1}$). For instance, the spectral density (with respect to wave-number magnitude) of kinetic energy, $E(\kappa)$ say, in parts of wave-number space corresponding to the equilibrium range is such a mean value and comes within the scope of the theory. Dimensional arguments show that the dependence on ϵ and ν requires

$$E(\kappa) = \epsilon^{1/4}\nu^{5/4}F(\kappa\nu^{3/4}/\epsilon^{1/4}), \qquad \ldots\ldots(4.3)$$

where F is a function of universal (and undetermined) form. The form (4.3) applies for values of κ satisfying $\kappa \gg 1/L$, irrespective of the large-scale properties of the turbulence, provided that the condition $\eta \ll L$ is satisfied. We note that the length η which measures the scale of the smallest existing eddies must be of

the order of magnitude of $(\nu^3/\epsilon)^{1/4}$, since no other characteristic length can exist in the equilibrium range. Since η is defined to order of magnitude only we can write

$$\eta = (\nu^3/\epsilon)^{1/4}, \qquad \ldots\ldots (4.4)$$

so that (4.3) becomes

$$E(\kappa) = \epsilon^{1/4}\nu^{5/4}F(\kappa\eta), \qquad \ldots\ldots (4.5)$$

provided $\kappa \gg 1/L$ and provided the Reynolds number is such that the condition

$$L\eta = L\epsilon^{1/4}\nu^{-3/4} \gg 1 \qquad \ldots\ldots (4.6)$$

is satisfied.

The prediction which can be made about $E(\kappa)$ when the Reynolds number is even higher is much more specific. Provided the Reynolds number is such that wave numbers satisfying (4.2) can be found, then for these wave numbers the expression (4.3) for $E(\kappa)$ must be independent of ν. This requires the unknown function $F(x)$ to be proportional to $x^{-5/3}$, giving

$$E(\kappa) = \alpha\epsilon^{2/3}\kappa^{-5/3} \qquad \ldots\ldots (4.7)$$

within the inertial subrange, where α is an absolute constant. At larger wave numbers, where viscosity affects the motion, the energy spectrum presumably falls off more rapidly. (We note in passing that the spectral density of mean-square vorticity, viz. $2\kappa^2 E(\kappa)$, increases with κ within the inertial subrange, as previously assumed.)

(ii) *Experimental Evidence*

Experimental tests of the above and similar predictions are not an easy matter since it turns out that the requirement (4.6) (and *a fortiori* (4.2)) cannot readily be satisfied with conventional laboratory equipment. Measurements of the turbulence produced by placing a uniform array of bars across a uniform stream in a wind tunnel ('grid' turbulence) show that the size of the grid and the velocity of the stream have to be inconveniently large if all the viscous dissipation is to occur at wave numbers large compared with $1/L$. For this reason conclusive experimental verification of the universal equilibrium theory has not yet been obtained, although all the evidence, so far as it goes, is consistent with the theoretical predictions. There is space here to record only three of the relevant measurements.

The prediction that the motion associated with the equilibrium range of wave numbers is always statistically isotropic has been tested by several experimenters (Townsend 1948, Corrsin 1949, Laufer 1950 a). Laufer used the turbulent flow of fluid produced by a pressure gradient in the space between two parallel planes, which has large-scale properties which are distinctly non-isotropic. By making experimental Fourier analyses of the variations of two components of the velocity recorded at a fixed point in the channel, he was able to compare the values of $\overline{B_1(\kappa)B_1^*(\kappa)}$ and $\overline{B_1(\kappa)B_2^*(\kappa)}$ for various values of κ (a suggestion due to Corrsin), as shown in fig. 1, where $\mathbf{B}(\kappa)$ is the Fourier coefficient concerned and suffixes 1 and 2 denote components parallel to the mean flow and perpendicular to the planes respectively. When the Fourier coefficient $\mathbf{B}(\kappa)$ is statistically isotropic, there is no correlation between $B_1(\kappa)$ and $B_2(\kappa)$, and fig. 1 shows that there is in fact a trend towards this state as κ increases.

Stewart and Townsend (1951) have made measurements of the one-dimensional spectrum $\phi(\kappa)$ (obtained from a Fourier analysis of the temporal variation of velocity at a fixed point) in grid turbulence, to which a prediction of the same

form as (4.5) should apply. Fig. 2 shows the results for various positions downstream from the grid and for various Reynolds numbers, plotted to exhibit the predicted similarity of shape under different conditions. The curves corresponding to different Reynolds numbers come towards coincidence as κ increases,

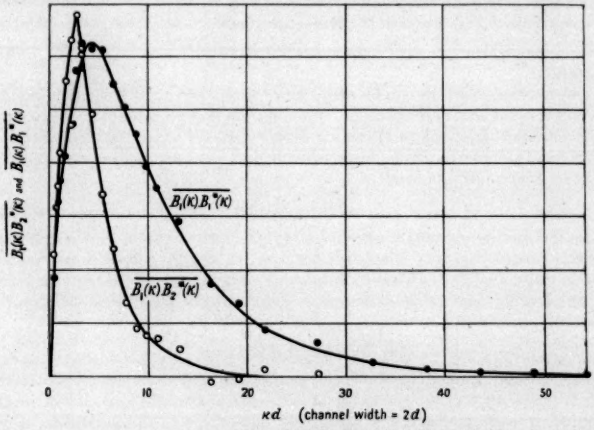


Fig. 1. Correlation between two components of the Fourier coefficient for flow in a channel. (*After* Laufer 1950 a)

Note. In the above, for $\overline{B_1(\kappa)B_1^*(\kappa)}$ *read* $\kappa^3\overline{B_1(\kappa)B_1^*(\kappa)}$
and for $\overline{B_1(\kappa)B_2^*(\kappa)}$ *read* $\kappa^3\overline{B_1(\kappa)B_2^*(\kappa)}$

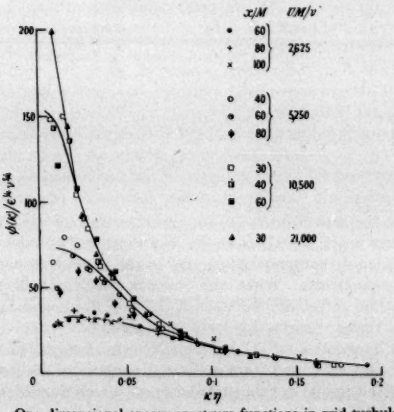


Fig. 2. One-dimensional energy spectrum functions in grid turbulence.

as they should according to (4.5). The coincidence is not quite as good as it appears to be from this figure, owing to the compression of the ordinate scale, and systematic differences between the curves can be detected for values of κ up to $\kappa\eta = 0{\cdot}6$. However, the coincidence is then within the experimental scatter, as is shown in fig. 3.

In both of the above tests of the equilibrium theory the Reynolds number was not so large that all the dissipation occurred at wave numbers large compared with $1/L$, and the agreement exhibited in figs. 1, 2 and 3 is not an immediate confirmation of the theory in the above form. Explanations of the agreement have been put forward, but the matter is too complex to be entered into here. The essential point seems to be that the restriction (4.6), which was introduced

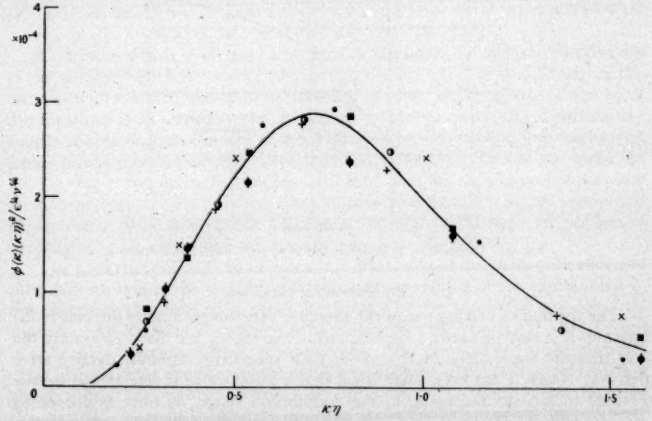


Fig. 3. One-dimensional energy spectrum function at large wave numbers.

in order to permit the deduction that the motion associated with the equilibrium range is determined, so far as the effect of the remainder of the turbulence is concerned, by ϵ, is necessary if the simplest possible conditions are to be achieved, but that, if it is not satisfied, the Fourier coefficients of large wave numbers nevertheless possess certain simple statistical properties which continue to be determined by ϵ in the special case of decaying homogeneous turbulence.

A third test is provided by measurements of dimensionless ratios of quantities which are determined by the equilibrium range of wave numbers. One such ratio is

$$\overline{\left(\frac{\partial v_1}{\partial x_1}\right)^2}\,\overline{\left(\frac{\partial^3 v_1}{\partial x_1^3}\right)^2} \Big/ \left[\overline{\left(\frac{\partial^2 v_1}{\partial x_1^2}\right)^2}\right]^2, \qquad \text{.....(4.8)}$$

where v_1 and x_1 are parallel velocity and position components. The differentials are equivalent to weighted integrals of the corresponding Fourier coefficients, and since the weighting factors are at least as high as in the expression for the viscous dissipation (which is proportional to a mean-square first derivative) the expression (4.8) is determined by the equilibrium range provided the condition

(4.6) is satisfied. Being dimensionless, and uniquely determined by ϵ and ν, it must be a universal constant. Measurements of the ratio in two different kinds of turbulence have been recorded (Batchelor and Townsend 1949) and are reproduced in fig. 4 (in both cases v_1 is the component of the velocity in the direction of the mean flow, and the abscissa is an effective Reynolds number suitable for comparison of the two turbulent motions). The measurements seem to show a trend towards a constant value near 4·5, and are consistent with the theoretical prediction, but higher Reynolds numbers are clearly needed for a more decisive test.

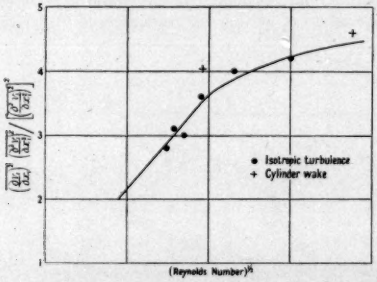


Fig. 4. Variation of a dimensionless ratio with Reynolds number.

(iii) *Applications of the Theory*

The method of making use of the theory is very simple; more difficult is the question if it *may* be used. Applications of the theory are made in exactly the way that the expressions (4.5) and (4.7) for the energy spectral density were derived. Once it has been established that a mean value is determined by the Fourier coefficients belonging to the equilibrium range, its form is chosen to conform with the requirement of isotropy (which can be done without difficulty, using the procedure set out systematically by Robertson (1940)) and with the dimensional requirements of dependence on the parameters ϵ and ν alone. Dimensional analysis is usually the only technique required for applications of the theory.

The ideal method of establishing if a certain mean value is determined by the equilibrium range is to express this mean value in terms of the Fourier coefficients. This expression will very often be an integral of some function of the Fourier coefficients over all wave-number space, and the criterion for determination by the equilibrium range is that the weighting at large values of κ should be strong enough to make the integral converge at $\kappa = 0$ when the Fourier coefficients in the integrand are given the values appropriate to the equilibrium range. However, it is not always possible to express the mean value concerned in terms of Fourier coefficients of the velocity distribution, and it may be necessary to use physical insight to decide if the mean value is influenced strongly by the small eddies.

Many interesting applications of the theory have already been made, and more will undoubtedly be made when it is more widely understood. The forms of many statistical quantities describing the velocity distribution (the energy spectrum being the simplest such quantity) have been predicted. Kolmogoroff (1949)

has used the theory to obtain information about the size of drops of liquid which exist in a fluid in turbulent motion under the opposing actions of surface tension and the shearing produced by the turbulence. Megaw (1950) has considered the scattering of light waves and short radio waves produced by variations of refractive index in the atmosphere, and, with the assumption that these variations arise from atmospheric turbulence, has used the theory to provide information about the high wave-number components of the distribution of refractive index. Another application relevant to atmospheric turbulence is to the prediction of the mean rate of separation of pairs of marked particles (e.g. small puffs of smoke) (Batchelor 1950 a). Other applications will be mentioned in the sections which follow.

§ 5. THE EFFECT OF TURBULENT MOTION ON THE DISTRIBUTION OF LOCAL PROPERTIES OF THE FLUID

A problem which is of special interest to physicists in view of the diversity of the fields in which it occurs is that of finding the effect of turbulence on the distribution (assumed non-uniform initially) of some local property of the fluid. Temperature is an obvious example of such a local property, and a particularly simple one, since each element of the fluid preserves its temperature unchanged (apart from the action of molecular conductivity), and the effect of the turbulent motion is to redistribute the temperature field. Despite the physical importance of such problems, they have not been studied widely, and it is not possible to report many definite advances. However, interest in the type of problem is quickening, and several lines of thought have been started.

The principal difficulty to be faced in a discussion of the convective effects of turbulent motion is the need for a Lagrangian description of the motion, that is, one which gives information about the history of the motion of any one fluid particle. An Eulerian description, which considers the velocity at a fixed point of space as a function of time, is considerably more convenient for the purposes of an analysis of the dynamics of the motion, and almost all the existing information about turbulence, either analytical or empirical, is in terms of an Eulerian description. Very little progress has been made towards relating the two ways of describing the turbulent motion, so that we are obliged to describe the convective effects of turbulence with the aid of quantities which are different from those used in most other work on turbulence.

Most of the physical properties whose spatial distributions are to be studied will be either scalar quantities that satisfy conservation of the total amount associated with a material volume* of fluid (e.g. temperature, concentration of suspended matter) or vector quantities which satisfy conservation of the total flux of the quantity through an open material surface of fluid (e.g. angular velocity or vorticity), and since it also happens that these are the fundamental cases from the point of view of the convection, the discussion will be restricted to these two kinds of local quantity. In both cases the effect of molecular conduction, which we have ignored for the purposes of the above classification, will be superimposed on the convective effects. If θ and $\mathbf{F}$ represent scalar and vector quantities of the above kind, the differential equations which they satisfy, in the absence of molecular diffusion, are

$$D\theta/Dt = 0, \qquad \ldots\ldots(5.1)$$

$$D(\mathbf{F} \,.\, d\mathbf{S})/Dt = 0, \qquad \ldots\ldots(5.2)$$

* A material volume of fluid is a volume which always contains the same particles of fluid.

where $D/Dt = \partial/\partial t + \mathbf{v}.\nabla$ is the operator giving the temporal rate of change as the fluid particle to which θ or $\mathbf{F}$ refers is followed in its motion, and $d\mathbf{S}$ is a (plane) element of a material surface.

Now in view of the incompressibility of the fluid the volume of an element of fluid remains constant as it follows the motion, so that if $d\mathbf{l}$ is a material (straight) line element which intersects $d\mathbf{S}$ we have

$$D(d\mathbf{l} \,.\, d\mathbf{S})/Dt = 0 \qquad \ldots\ldots (5.3)$$

Hence the meaning of (5.2) is that the value of $\mathbf{F}$ associated with a particle of fluid changes, as the particle follows the motion, in the same way as a material line element, both in magnitude and direction. That is, lines of $\mathbf{F}$ follow the motion of the fluid, and the local magnitude of $\mathbf{F}$ is proportional to the length of a material line element coincident (locally) with the line of $\mathbf{F}$. The problem of determining the convective changes in the distribution of $\mathbf{F}$ is therefore identical with a consideration of the history of lines moving with the fluid. On the other hand, level surfaces of θ move with the fluid, in view of (5.1), and the determination of the convective changes in the distribution of θ is identical with a consideration of the history of surfaces moving with the fluid. This is the sense in which the local quantities represented by θ and $\mathbf{F}$ have fundamental convective properties.

The qualitative effects of the turbulent motion can now be determined. One of the basic properties of turbulent motion is its diffusive character; the difference between the velocities of two fluid particles fluctuates randomly with time, and the particles, if initially close, necessarily tend to move apart. It may be that the boundaries prevent the distance between the two particles from tending to increase indefinitely*, but *neighbouring* particles always tend to separate. Consequently line elements moving with the fluid always tend to lengthen (and to lengthen without limit, irrespective of the presence of boundaries, since, however long the line may become, each element of it continues to tend to lengthen). The implication of this is that the magnitude of $\mathbf{F}$ tends to increase in all parts of the fluid as the lines of $\mathbf{F}$ are distorted and extended by the turbulence. Quantities like $\mathbf{F}^2$ will thus increase with time. There will also be changes in the shape of the distribution of $\mathbf{F}$, but these are less readily analysed since they depend on the spectrum of the turbulence and on the initial spectrum of the spatial distribution of $\mathbf{F}$. The distortion of the lines of $\mathbf{F}$ is a non-linear interaction process, which is described by the equation, readily derived from (5.2),

$$D\mathbf{F}/Dt - \mathbf{F} \,.\, \nabla\mathbf{v} = 0, \qquad \ldots\ldots (5.4)$$

and is comparable in complexity with the process of inertial transfer of energy from one Fourier component (of the velocity distribution) to another. However, it is at least clear that the changes in the spectrum of $\mathbf{F}$ will be most important in the neighbourhood of wave numbers at which the lengthening effect due to the turbulence is large.

The general effect of the convection on the distribution of θ can be determined by the same arguments. We can see that just as material line elements tend to lengthen in turbulent motion, so neighbouring surfaces tend to come closer together as they follow the motion. For consider a small parallelepiped defined by the three intersecting material line elements $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, and allow this volume element to follow the motion. Each of the three line elements tends to lengthen,

* In an unlimited fluid the mean square of the distance apart would ultimately increase linearly with time (Batchelor 1949 b) as in other ' random walk ' problems.

and in order to maintain a constant volume ($=\mathbf{a}.\mathbf{b}\times\mathbf{c}$) the angles between the line elements must tend to decrease. The volume is a product of three lengths and two angles (when the latter are small), so that the relative rate of decrease of the angles must be faster than the relative rate of increase of the lengths. Hence the perpendicular distance between two opposite faces of the parallelepiped—which can be considered as elements of two neighbouring non-intersecting material surfaces—tends to decrease. Level surfaces of θ thus tend to come closer together under the diffusive action of the turbulence. As a corollary we see that the gradient of θ tends to increase at all points of the fluid.

An estimate of the rates at which line elements tend to lengthen and neighbouring surfaces tend to come together is provided by the universal equilibrium theory in cases in which the contributions from the distribution of mean velocity may be ignored. The rate of extension of a line element is a spatial gradient of velocity, and it is a reasonable assumption that when the Reynolds number is sufficiently large it is determined by the 'small' eddies, i.e. by Fourier coefficients belonging to the equilibrium range of wave numbers. We can therefore predict that statistical quantities like the mean relative rate of extension $\overline{(1/dl)(\partial dl/\partial t)}$ are determined by the (local) parameters ϵ and ν together with any parameters relating to the initial selection of the material line element of length dl. At the instant of selection of the line element $\overline{(1/dl)(\partial dl/\partial t)}$ is necessarily zero, since the mean velocity at each end of the line element is zero, and thereafter its value increases. At a sufficient time after the selection of the line element the influence of the initial direction of the line element disappears and $\overline{(1/dl)(\partial dl/\partial t)}$ presumably tends to a constant value. In this asymptotic stage (which is set up after a time interval small compared with times characteristic of the turbulence as a whole) we have, according to the universal equilibrium theory,

$$\overline{(1/dl)(\partial\, dl/\partial t)} = c\,(\epsilon/\nu)^{1/2}, \qquad \ldots\ldots(5.5)$$

where c is an absolute constant. Likewise the mean relative rate of decrease of the perpendicular distance between neighbouring surfaces tends to a constant value which is found, from the isotropy of the small-scale stretching process, to be a half of the expression (5.5).

As already mentioned, molecular conduction or diffusion will usually be superimposed on the convection caused by the turbulence. Taking both effects into account, the differential equations for θ and $\mathbf{F}$ are readily found to be

$$D\theta/Dt = k\nabla^2\theta, \qquad \ldots\ldots(5.6)$$

$$D\mathbf{F}/Dt = \mathbf{F}.\nabla\mathbf{v} + k\nabla^2\mathbf{F}, \qquad \ldots\ldots(5.7)$$

where k is an appropriate molecular diffusivity. The effect of the molecular conduction is to smooth out the variations of θ and $\mathbf{F}$, so that the convection and the conduction have opposing effects on the vectors $\nabla\theta$ and $\mathbf{F}$. Clearly, if the diffusivity k is small enough, quantities like $\overline{(\nabla\theta)^2}$ and $\overline{\mathbf{F}^2}$ will increase with time, whereas if k is large enough the damping effect will be stronger than the convective effect and they will decrease. It has been argued (Batchelor 1950b) that the critical value of k is of the order of magnitude of ν, on the basis that ν is the diffusivity for the particular choice of $\mathbf{F}$ (viz. vorticity) for which the convective and conductive effects are necessarily in equilibrium. Equations (5.6) and (5.7) are linear in θ and $\mathbf{F}$, so that if the value of k is below the critical the convective effect will be dominant irrespective of the magnitude of $\nabla\theta$ or $\mathbf{F}$ (although not

altogether irrespective of the spectrum of $\nabla\theta$ or $\mathbf{F}$); the fluid in turbulent motion will thus be unstable to the introduction of a small disturbing field of $\nabla\theta$ or $\mathbf{F}$ and, once introduced, these fields will grow in intensity. So far as the above equations are concerned they will grow without limit, but in reality there will usually be some back reaction which limits their magnitude.

The effect of turbulent motion on the distribution of various physical quantities has been studied in the relevant special contexts, but all the quantities considered in the literature are included within the above general argument (for further details of which see Batchelor 1952), and there is no need to rephrase the results for each physical quantity. Vorticity was the first quantity to be studied in the above Lagrangian manner (Taylor 1938 a), and some experimental results are also available for this quantity in the case of turbulence produced by a grid (Batchelor and Townsend 1947). Vorticity is one of the parameters describing the motion that causes the extension of the lines, so that the convection of vorticity and the convecting motion are more properly regarded as different aspects of the same motion. It has been shown (Batchelor 1950 b) that magnetic field in a conducting liquid is another vector quantity, the convection of which is described by equation (5.2). The diffusivity k is in this case equal to $(4\pi\mu\sigma)^{-1}$, where μ is the magnetic permeability, and σ the electrical conductivity, of the liquid. The value of $(4\pi\mu\sigma)^{-1}$ for liquid metals such as mercury is very far from satisfying the condition for a small magnetic field to be amplified by the turbulent convection but the condition may be satisfied in the case of rarefied highly ionized gases. If amplification does occur, the ultimate limit to the strength of the magnetic field is set by the stress which the field exerts on the medium; when the magnetic field is strong enough it suppresses (in part) the motion which is amplifying it. The case of temperature has also been studied (Obukhoff 1949 a, Yaglom 1949, Inoue 1950, Corrsin 1950), and interesting results have been obtained on the basis of plausible extensions of the universal equilibrium theory.

§6. HOMOGENEOUS DECAYING TURBULENCE

The remarks made in the preceding sections apply to all cases of turbulent motion without restriction (except that the Reynolds number was required to be large for the theory of §4 to apply), but it is necessary now to particularize the boundary and initial conditions to some extent. Some of the properties of turbulent motion are known to be markedly dependent on the nature of the boundary conditions, and since there is not yet available a comprehensive analysis of the problem we are obliged to examine the important cases separately. The case of homogeneous turbulence—that is, a motion whose statistical properties are independent of position and which is necessarily decaying in view of the absence of a source of energy of the fluctuations—is not very important technically, since in nearly all cases of flow through pipes, etc., the effect of the boundaries is to make the turbulence decidedly inhomogeneous. However, it is in some respects a dynamically simple case of turbulent motion and offers some possibility of theoretical progress.

The scientific importance of the case of homogeneous turbulence was greatly enhanced when Taylor established in 1935 that a close approximation to it could be generated in laboratory wind tunnels; this possibility of convenient experimental work and the parallel developments in the technique of measurements with the hot-wire anemometer (an account of which is given in the Appendix) have given a tremendous impetus to the whole subject. Taylor found that if

a regular array of bars spaced a distance M apart in each of two perpendicular directions is placed across a uniform wind-tunnel stream (of velocity U) the turbulent wakes behind the bars amalgamate at a short distance downstream, and at a distance of about $5M$ from the grid there is no detectable variation of the mean velocity across the stream. Further downstream, at a distance of about $10M$, the mean quantities describing the velocity fluctuations are found to be constant over the lateral plane, and this property is retained at greater distances from the grid. The turbulence becomes weaker with increasing distance from the grid, since it is being dissipated by viscosity, but it is found that the spatial variation of the mean quantities in the direction of the stream is so small that for all practical purposes the turbulence can be regarded as spatially homogeneous. Thus if x denotes distance from the grid, x/M is a non-dimensional parameter which specifies the stage of the decay process and which can be regarded equally as a non-dimensional decay time tU/M appropriate to the (mathematical) problem of exactly homogeneous turbulence.

The earliest experimenters also found that the mean squares of the three velocity components are approximately equal at distances not less than about $10M$ from the grid, and subsequent tests have established the more general conclusion that the turbulence generated by the grid rapidly becomes approximately isotropic. This additional simplicity of grid turbulence has been a source of great stimulus to mathematicians, who have been encouraged to investigate the consequences of isotropy for many of the complicated tensor statistical quantities needed to describe the turbulence. This work has developed from papers by von Kármán and Howarth (1938) and by Robertson (1940), and much formal analysis of great elegance has been published. However, it must be admitted that this analysis has been concerned chiefly with the kinematics of turbulence, and has not contributed much to our understanding of the dynamics of the problem.

From the dynamical point of view the feature of isotropic homogeneous turbulence which interests us most is the rate of decay of kinetic energy owing to viscous dissipation ($=\epsilon$). Viscous damping is a linear process which depends on the energy spectrum in a simple way—we see from (3.4) that if $E(\kappa)\,d\kappa$ is the contribution to the kinetic energy per unit mass of fluid from Fourier components with wave numbers whose magnitudes lie between κ and $\kappa+d\kappa$, the corresponding contribution to the rate of dissipation is $2\nu\kappa^2 E(\kappa)\,d\kappa$—and the real question is, what shape does the spectrum take as a consequence of the combined action of inertial transfer across the spectrum and viscous damping? In view of the fact that at large Reynolds numbers the bulk of the viscous dissipation occurs at high wave numbers where $E(\kappa)$ has fallen to a comparatively small value, the problem of determining the rate of decay can be regarded, in a sense, as concerning the non-linear inertia effect alone. The inertial self-modulation of the energy-containing range of wave numbers (which is centred on a wave number of the order of magnitude $1/M$) determines the rate ϵ at which energy is transferred to larger wave numbers which, as we saw in § 4, are then required to adjust themselves to dissipate energy at this rate. Clearly the determination of the rate of decay is a difficult problem, in general only likely to be soluble in specially simple circumstances.

The measured law of decay of turbulence generated by a grid is invitingly simple in form. Many measurements (Batchelor and Townsend 1948a, Stewart and Townsend 1951), at different values of the Reynolds number UM/ν

and with grids constructed in different ways, have shown that the energy decays according to the law

$$\frac{U^2}{\overline{v^2}} = \frac{b}{C_D}\left(\frac{x-x_0}{M}\right) \qquad \ldots\ldots(6.1)$$

for distances from the grid up to about $x = 150M$. C_D denotes the resistance coefficient for unit area of the grid, x_0 is the position of the virtual origin of the turbulence (which usually lies between $5M$ and $15M$), and b is a constant depending partially on the shape of the grid (having the value 100 for a square-mesh grid made of circular bars of diameter $M/5{\cdot}33$). Fig. 5 shows a representative set of measurements of the energy of the turbulence in the period of decay—termed the 'initial period'—for which (6.1) holds. At larger values of x/M the energy

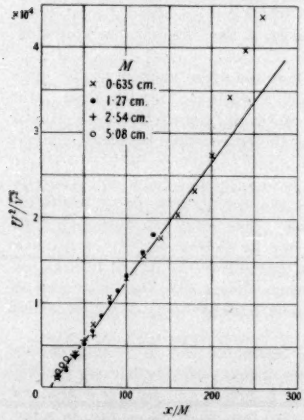


Fig. 5. Decay of energy during the initial period ($U = 1\,286$ cm.sec^{-1}).

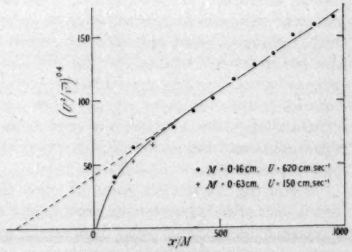


Fig. 6. Decay of turbulent energy in the final period.

begins to decay more rapidly than (6.1) would predict and ultimately (after a non-dimensional time of decay which increases rapidly with increase of the Reynolds number) the decay proceeds according to the law

$$\frac{U^2}{\overline{v^2}} \propto \left(\frac{x-x_1}{M}\right)^{5/2}, \qquad \ldots\ldots(6.2)$$

where x_1 is another virtual origin. Some measurements of the energy of one velocity component (in the direction of the stream) in this final period of decay (taken from Batchelor and Townsend 1948 b) are shown in fig. 6; the agreement with the relation (6.2) is not compelling, but this law has also been established from rigorous mathematical arguments.

Before considering the more difficult problem presented by the initial period of decay, we can dispose of the final period. It is clear that the ultimate effect of the decay must be to produce a turbulent motion so weak that the non-linear terms in the Navier–Stokes equation (2.2) are negligible. When this stage is reached, the transfer of energy across the spectrum ceases and each Fourier

component of the velocity distribution decays independently. Since the relative rate of dissipation of a Fourier component is proportional to κ, the Fourier components which survive longest are those with wave-number magnitudes in the neighbourhood of $\kappa=0$. Now it can readily be shown that the expansion of the energy spectrum function $E(\kappa)$ as a power series in κ^2 begins with the term in κ^4; the coefficient of κ^0 is of course zero since $E(\kappa)$ is an energy density with respect to wave-number magnitude, and the coefficient of κ^2 is found to be zero as a requirement of the equation of continuity. Hence the ultimate shape of the spectrum is that which is obtained when an initial spectrum $E(\kappa)=C\kappa^4$ (C = constant) decays linearly, under the action of viscosity alone; that is, the ultimate shape is

$$E(\kappa)=C\kappa^4\exp(-2\nu\kappa^2 t). \qquad \ldots\ldots (6.3)$$

It follows at once that the total energy $\int_0^\infty E(\kappa)\,d\kappa$ varies as $t^{-5/2}$, as indeed is found experimentally at a sufficiently advanced stage of the decay. The spectrum shape (6.3) has also been confirmed by measurements in the final period of decay.

One further development may be mentioned in connection with the final period of decay. It can be shown (Batchelor 1949a) that the coefficient C is independent of time, not only in the final period (which is evident) but also in the initial period and throughout the whole history of the decay; the biggest eddies experience neither viscous damping nor inertial modulation on account of the very small spatial gradients of velocity associated with them. The value of C obtained from measurements of the decay of energy in the final period is thus a piece of information about the way in which the field of homogeneous turbulence is generated by the grid. These remarks apply not only to the coefficient C appropriate to total energy, but also to the various coefficients needed to describe the energy spectrum tensor at small values of κ. It follows that unless the homogeneous turbulence is generated isotropically (which is unlikely) the components with very small wave numbers will never become isotropic. The lack of isotropy of the small wave-number components generated by a grid does in fact show up quite clearly in the final period of decay (Batchelor and Stewart 1950), even though it is not detectable in the initial period.

Returning now to the initial period of decay, the noteworthy observation is the extent to which the turbulence rapidly settles down to a state which is approximately independent of the conditions of its formation. For instance, independently of the Reynolds number, and for the several different shapes of grid that have been used, the turbulence quickly becomes (approximately) isotropic and the total energy decreases according to the general law (6.1). More detailed evidence on this point is provided by measurements of the spectrum and correlation functions. Stewart and Townsend (1951) have assembled a large amount of evidence, which cannot be summarized adequately here, and in fig. 7 is reproduced one of their sets of measurements of the spectrum function $E(\kappa)$ at different times of decay during the initial period for a square-mesh grid (actually velocity correlation functions were measured and spectrum functions computed from them using the well-known Fourier transform relations). The curves are normalized to have the same area, and the length λ used as a unit is given by

$$\lambda^2=10\nu(t-t_0)=(10\nu/U)(x-x_0). \qquad \ldots\ldots (6.4)$$

The figure shows that soon after its generation by the grid the turbulence adjusts itself so that the subsequent changes are principally in the velocity scale (a

decrease as $(x-x_0)^{-1/2}$) and in the length scale (an increase as $(x-x_0)^{1/2}$). This partial preservation of the shape of the spectrum during decay, together with evidence that the spectra at other Reynolds numbers and for other grids coincide approximately with the curves of fig. 7, especially at the higher wave numbers, is strong evidence for the existence of some kind of asymptotic statistical state. However, the variations in the shape of the spectra at the smaller wave numbers in fig. 7 are real and affect a proportion of the energy that is not negligible. How it comes about that these significant changes of shape of the spectrum can occur during decay while at the same time the total area under the spectrum curve decreases according to the simple law (6.1) is not at all clear.

Attempts to explain the observed law of decay have been made by extending the ideas of the universal equilibrium theory. If there is a range of (large) wave numbers for which the associated Fourier components are in approximate

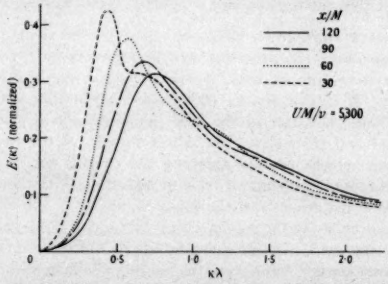


Fig. 7. Three-dimensional spectrum functions at different times of decay.

statistical equilibrium (in virtue of the fact that their characteristic times are small compared with the decay time), it is reasonable to suppose that the Fourier components at slightly smaller wave numbers are in a kind of quasi-equilibrium which is not uninfluenced by the existence of the decay and which will require at least one parameter additional to ϵ and ν for its statistical specification. Heisenberg (1948 b) postulated that most of the energy of the turbulence is included within this quasi-equilibrium, and if t is chosen as the additional parameter needed to determine the motion statistically, a decay law of the form (6.1) follows immediately from the need for $\epsilon t^2/\nu$ to remain constant. However, such a postulate implies similarity of shape of almost the whole of the spectrum during decay, contrary to evidence of the kind shown in fig. 7. Goldstein (1951) and von Kármán and Lin (1951) have suggested that the range of wave numbers over which the quasi-equilibrium operates is more restricted and does not include most of the energy, which avoids the difficulty of predicting too much about the behaviour of the spectrum; however, the physical basis for a prediction about the decay of total energy is then less clear.

§7. FLOWS IN WHICH THERE IS A MEAN SHEARING MOTION

In all cases of turbulent motion in which the presence of boundaries influences the motion, and also in certain other cases, the mean velocity varies with position. We can regard the turbulent fluctuations in such cases as being generated by the mean shearing motion which in turn is (usually) generated by motion of the

boundaries. In spite of the fact that the turbulence associated with a mean shearing motion is frequently statistically steady, and may even be homogeneous with respect to one or two position coordinates (as, for example, in the case of flow through a pipe under pressure), the analytical difficulties caused by the necessary inhomogeneity with respect to at least one of the position coordinates are very severe. Spatial non-uniformity is more difficult to handle than temporal non-uniformity, since the quadratic term of the Navier–Stokes equation involves spatial gradients, and it is necessary to take account of transport of quantities such as the turbulent energy from one point to another. Historically, such problems as the determination of the distribution of mean velocity in cases of turbulent shear flow were the first to be studied (by Reynolds, Boussinesq, and others), but this was because of their obvious technical importance; they may well be the last to be solved.

At the present time the picture which we have of the mechanics of turbulence in shear flow is clear in certain limited respects only, and a unified treatment of the problem is not available. There would not be space here to describe the several experimental investigations of particular shear-flow systems that have been made, and we shall consider only the very general properties of such systems. For the most part these properties have to be described in qualitative terms, although some of the current work is directed towards providing an empirical analytical description of the motion.

During the ten years ending about 1946 the principal advance was the growing conviction that the mixing-length type of theory was inadequate, not merely in the details of its predictions but also in its basic postulates. For so long as the theories of this kind were used to predict no more than the distribution of mean velocity no serious errors were found, but this seems to be due, in part at least, to the relatively small permissible variation of the velocity distribution when the boundary and symmetry conditions have been satisfied. More sensitive tests of the mixing-length theories, such as the prediction of the distribution of turbulent energy, have revealed their weakness. The essential error seems to lie in the assumption that turbulent transport in the direction of the spatial gradient of some mean quantity is a small-scale mixing process, or 'gradient' type of diffusion, which can be specified by a turbulent transport coefficient. The various mixing-length theories (see *Modern Developments in Fluid Dynamics*, Oxford University Press, 1938) differ among themselves as to how this turbulent transport coefficient should be evaluated, but they have the above assumption in common, and they all assume, as a corollary, that the transport coefficient at any point is determined by the local conditions.

It is not necessary to retrace the long and rather confusing history of the arguments and counter-arguments which have led to a disbelief in the (undoubtedly simple) notion of a local transport coefficient, since one or two inconsistencies can be seen immediately. The 'mixing length', or equivalent mean free path, of the transport process can be calculated from observations of the rate of transport, using the formulae supplied by the above theories, and it is almost always found that the mixing length is not small by comparison with lengths representative of the distribution of the quantity being transported. Likewise the mixing time is usually found not to be small compared with the time which a fluid particle would take to move from one position to another at which the local mean properties are appreciably different. These internal inconsistencies of the assumption of a gradient type of diffusion, with a diffusion coefficient determined by local

conditions, have been made particularly evident from measurements in a two-dimensional half-jet (Liepmann and Laufer 1947) and in a two-dimensional wake (Townsend 1949 a). The conclusion, which is now widely accepted, is that the rate of transport of some mean quantity at any point in the fluid is determined, partially at least, by the field of motion as a whole, and cannot be related to local quantities. The phenomenon of turbulent shear flow is evidently more difficult than was originally thought.

Since 1946 research has been directed towards the more constructive end of making clear the elementary mechanical processes occurring in shear flow. Definitive experimental studies have been made of the two-dimensional half-jet (Liepmann and Laufer 1947), the round jet, with and without heating (Corrsin and Uberoi 1949, 1950), the two-dimensional wake (Townsend 1947 b, 1949 a, b, 1950), the two-dimensional channel (Laufer 1950 a, b), the boundary layer without pressure gradient (Townsend 1951 a), and the boundary layer with pressure gradient (Dryden 1948, Schubauer and Klebanoff 1950). In each case many quantities other than mean velocity, e.g. energy spectrum functions, turbulent shear stress, rate of transport of turbulent energy, were measured, and much useful information has been collected. It cannot be said that a proper understanding of shear flow turbulence in general has emerged from these studies, but it may do so when they have been given further thought.

It seems to be useful to distinguish between cases of turbulent flows which are closed in laterally by rigid boundaries (e.g. flow in pipes and channels, flow between rotating concentric cylinders, etc.) on the one hand, and cases of free turbulent flow which are surrounded by stationary fluid (e.g. wakes and jets) on the other. It has long been appreciated that these two types of shear flow would have different properties—for instance, the momentum and vorticity transfer theories were at one time believed to be applicable to cases of closed and free turbulent flows respectively—but the effects of the juxtaposition of fully developed turbulent motion and stationary fluid in cases of free turbulent flow have been understood only recently as a result of Townsend's work on the wake behind a long cylinder. Townsend has been led to infer from measurements of many statistical quantities that the wake consists at any moment of a core of small-scale intense eddying motion divided from almost stationary fluid by a well-defined boundary; the boundary is irregular and has large-scale protuberances which move outward and increase the width of the core as the wake develops. The most direct confirmation of this wake structure is provided by pictures of the wake at one instant, of which fig. 8 (Plate) shows a sample for the case of a bullet moving at supersonic speed. The uniformity of the texture of the turbulent motion within the core which is evident from the photographs is not only apparent but can also be inferred from measurements. The wake core increases in width with distance downstream by entraining non-turbulent fluid across the boundary, and the motion inside the core is so rapid that the entrained air is brought to the same state of small-scale motion almost immediately.

An immediate consequence of the irregularly shaped sharply defined boundary of the wake core is the phenomenon of intermittency of the fluctuations recorded by a measuring instrument at a fixed point not too close to the centre of the wake. For instance, a hot-wire anemometer shows a trace on a cathode-ray oscillograph which fluctuates for a time in the normal way and then has a period of slow low-amplitude variation, the periods of activation and quiescence following each other

(*a*)

(*b*)

Fig. 8. The instantaneous structure of a wake: (*a*) close to the bullet, (*b*) about 120 diameters downstream.

To face page 124

in a random way. If γ denotes the probability that the measuring instrument is immersed in turbulent fluid, i.e. that it lies inside the instantaneous wake core, γ is found to vary continuously from unity at points near the centre of the wake to zero well outside the wake, in a way which presumably depends on the shape and number of the protuberances at the edge of the wake core. One method of determining γ which is found to give consistent results is based on the assumption that the probability distributions of the velocity component parallel to the mean flow and of the various velocity derivatives are the same at all points within the wake core, the measured variation across the wake being due wholly to the variation in the total time for which the measuring instrument is immersed in turbulent fluid. The distribution of γ across the wake obtained in this way by Townsend is compared in fig. 9 with the distribution of mean velocity at a typical cross section of the wake. Clearly the effect of the intermittency will not be confined to the outer edge of the wake.

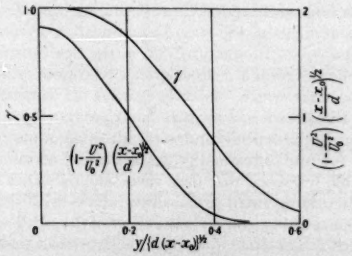


Fig. 9. Distribution of mean velocity U and intermittency factor γ across the wake of a cylinder (d = diameter of cylinder, x = distance from cylinder, $x_0 = 60\,d$).

With this picture of the structure of the wake in mind it is possible to give a description of the mechanism of turbulent transport across the wake which is consistent with the available measurements. It appears that there are two distinct ways by which quantities may be transported; first, the vigorous small-scale motion within the core can act as an interchange mechanism, giving a gradient type of diffusion for the fraction γ of the time for which it operates, and second, the outward velocity of the protuberances at the edge of the wake core can convect outwards any quantities, such as energy of the small-scale motion, which are finite within the core and zero outside. The diffusion mechanism applies to all quantities, whereas the convection mechanism applies only to those quantities associated with the core and presumably does not apply to momentum of the mean flow in a simple way, since this is not a local property of fluid. Townsend (1949 a) has shown, with the simple assumption that mean momentum is transported wholly by the small-scale mixing, that the effective eddy viscosity within the wake is $\eta = \overline{uv}/\gamma\partial U/\partial y$, where u, v are turbulent velocity components parallel to and across the mean flow and U is the mean velocity, and that η is approximately constant across the wake; this is consistent with the hypothesis that the small-scale motion within the core is approximately statistically uniform across any section.

The contribution to the transport from the small-scale mixing should be described by the same diffusion coefficient η for all mean quantities. In the case of quantities such as temperature (when the cylinder generating the wake is heated) and mass concentration of some second fluid (in the case of a jet which is initially composed wholly of this second fluid) there will be an additional contribution to the lateral transport from the outward convection at the edge of the core. It is difficult in such cases to separate the two contributions to the transport, but Townsend has come to the conclusion that in general they are of comparable importance. This is evidently the explanation of the observation (Hinze and van der Hegge Zijnen 1949, Corrsin and Uberoi 1949) that in free turbulent flows heat and matter spread at an equal rate and more rapidly than momentum.

To complete the qualitative description of the mechanics of free turbulent flows it is necessary to make clear the controlling mechanism which determines the relation between the intensity of the small-scale motion inside the core and the intensity of the large-scale motion manifested as protuberances at the edge of the core. Measurements of the (one-dimensional) spectrum functions of the three velocity components (Townsend 1950) in the two-dimensional wake show that the large-scale eddies form a distinct group, with linear dimensions comparable with a half the wake width, and suggest that the vorticity associated with them lies chiefly in the direction of the maximum positive mean rate-of-strain, i.e. at 45° to the direction of the cylinder and to the direction of mean flow. They are favourably aligned for intensification by the stretching action of the mean flow, and it is a reasonable inference that they came into existence as a result of the stretching process acting on small casual disturbances with the right alignment. In opposition to the stretching process is the effect of the small-scale motion which acts as an eddy viscosity and tends to weaken the large-scale eddies. Townsend's idea is that if the stretching process leads to a particularly strong local large-scale eddy, the increased rate of convection and rate of entrainment of non-turbulent air at the edge of the core will in turn lead to an increased eddy viscosity due to the small-scale motion and thence to a reduction in the intensity of the large-scale motion. There is thus the possibility that the two effects are approximately in an equilibrium which determines the relation between the intensities of the small and the large-scale motions, and Townsend has made this hypothesis plausible by considering what would happen to large-scale eddies with a prescribed simple velocity distribution.

Cases of closed turbulent flows have not been studied as fully as the two-dimensional wake, and it would be premature to attempt to interpret the observations even qualitatively. It has at least been made clear that the boundaries play a dominant part in the flow, and that nothing like the intermittency phenomenon occurs in this case. Both Laufer's (1950 a, b) observations of flow in a two-dimensional channel and Townsend's (1951 a) observations of a boundary layer (which is a mixed case, inasmuch as the motion in the outer part of the layer resembles a free turbulent flow, but is quite different nearer the wall) show that the viscous dissipation of turbulent energy is strongly localized to a thin layer close to the boundary. The turbulent energy flows across the layer towards the wall, chiefly through the action of a correlation between the fluctuating velocity normal to the wall and the strong fluctuating pressures which exist in the region where there is a high rate of mean shear. The velocity fluctuations are quite a large fraction of the local mean velocity in this dissipation region; fig. 10 shows,

for example, the root mean square of the (fractional) velocity fluctuation parallel to the mean flow in the case of the two-dimensional channel (Laufer 1950 a). The laminar sublayer, defined as the region in which viscous forces are dominant and the mean velocity profile is linear, extends beyond the maximum of $(\overline{u^2})^{1/2}/U$ in each of the three cases described in fig. 10.

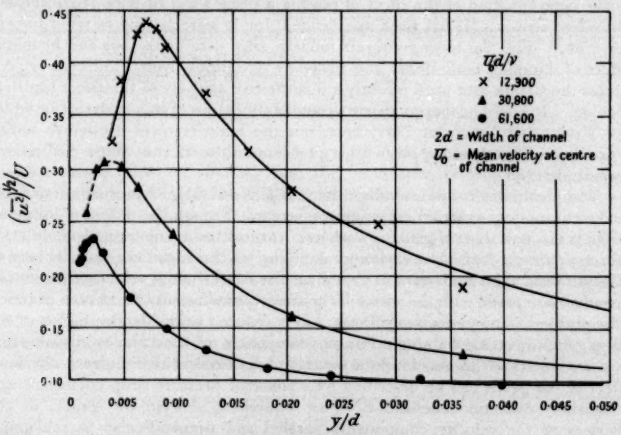


Fig. 10. Longitudinal velocity fluctuations near a wall.
(*After* Laufer 1950 a)

In both the channel and the boundary layer the mean velocity was found to be approximately proportional to the logarithm of the distance from the wall, for points neither too close to nor too far from the wall. This well-known logarithmic distribution has been used to support many different theories over the last twenty years, most of them being based on the assumption that the exchange of momentum can be described by an eddy viscosity coefficient determined by local conditions. In opposition to this assumption is the calculation, from the measured turbulent shear-stress, that the required mixing paths and mixing times are not small by comparison with lengths and times representative of the mean velocity distributions; in other words, eddies of a size comparable with the distance from the wall make a significant contribution to the shear stress. The explanation of the logarithmic distribution of mean velocity—and, indeed, of several similar laws for other flows—is probably to be found in simple considerations of similarity, as has been emphasized by Squire (1948). It can be argued that in the neighbourhood of the wall (although outside the thin laminar sublayer) the only length relevant to the flow is y (= distance from wall) and the only relevant velocity is the friction velocity $U_*(=(\tau/\rho)^{1/2}$, where τ is the shear stress at the wall), so that the mean vorticity must be of the form $\partial U/\partial y = \text{const.}\,(U_*/y)$, whence the logarithmic distribution follows. It is not possible to apply this similarity argument to the function $U(y)$ directly, since the velocity relative to the wall is not independent of the laminar sublayer and hence depends on the viscosity as well as on y and U_*.

§8. MISCELLANEOUS DEVELOPMENTS

(i) *The Passage of Homogeneous Turbulence through Gauze*

A practical problem of some interest in connection with the design of wind tunnels and water channels in which a small intensity of turbulence is required is the determination of the effect of placing a plane sheet of wire gauze across a turbulent stream. It has been appreciated, for at least as long as wind tunnels have been used for testing aircraft models, that wire gauze has the beneficial effect of damping both steady and unsteady deviations from a uniform velocity across the stream, but until recently a quantitative analysis of the effect has been lacking. It would be inappropriate to record the details of this analysis (for which see Taylor and Batchelor 1949) here, but the main features are worth noting since they will probably apply to other problems of the effect of some disturbance to turbulent flow.

The idealization which renders the problem tractable is to imagine that the turbulent motion is carried, by a uniform stream, through the region of influence of the transverse sheet of gauze in such a short time that negligible self-modulation (through inertia forces) and viscous damping of the turbulent motion occurs. This is found to be equivalent to assuming that the turbulent velocity fluctuations are small compared with the speed of the stream, as is usually the case in practice. The problem is now effectively linear, and is reduced to the determination of the effect of the gauze on a single Fourier component of the velocity distribution. It turns out that this can be done exactly if, as experiments suggest, the local effect of the gauze can be described by a constant pressure-drop coefficient and a constant deflection coefficient. The reductions, due to the gauze, in the energies of the velocity components parallel and perpendicular to the gauze (which are different: the gauze has a directional effect, and turbulence which is isotropic far upstream is made axially symmetric about the normal to the gauze) are then found by integration over all wave numbers. Townsend (1951 b) has obtained experimental confirmation of the theoretical analysis under conditions approximating to those assumed in the theory.

The effect of passing a turbulent stream through a sudden contraction or distortion of the stream boundaries can be analysed in very much the same way, as was first pointed out by Taylor (1935 b), although the approximation needed for linearization, viz. that no self-modulation of the turbulence occurs during the passage through the region of the distortion, is here much less likely to be satisfied in practice. However, both of these problems provide useful reminders that the advantages of linearization are no less considerable in problems of turbulent motion than in other problems of mechanics.

(ii) *Hypotheses about the Transfer of Energy across the Spectrum*

It has already been pointed out the that principal difficulty to be faced in an analysis of turbulent motion is that presented by the inertial interaction of the different Fourier components. This interaction leads to a statistical flow of energy between different parts of wave-number space, and has a direct influence on such important quantities as the shape of the energy spectrum and the rate of viscous dissipation. A mathematical determination of the inertial transfer of energy seems to be out of the question at present, on account of the difficulty of handling non-linear equations, and attention has therefore been directed toward

the establishment of physical hypotheses about the energy transfer which by-pass the mathematical difficulties. Several such hypotheses have been made, one of which deserves mention here.

The object which these hypotheses have is to establish an expression for the rate, $S(\kappa)$ say, at which energy is transferred from wave-number magnitudes lower, to those higher, than κ (the more general quantity describing the flow of energy from any part of wave-number space to any other is beyond the reach of physical intuition). The equation for the variation of the spectral density (with respect to wave-number magnitude) of energy, $E(\kappa)$, is then

$$B(\kappa) - \frac{\partial}{\partial t}\int_0^\kappa E(\kappa')\,d\kappa' = S(\kappa) + 2\nu\int_0^\kappa \kappa'^2 E(\kappa')\,d\kappa', \qquad \ldots\ldots(8.1)$$

where $B(\kappa)$ is the rate of supply of energy, to Fourier components with wave-number magnitudes less than κ, from the boundaries or any external source. If $S(\kappa)$ can be given in terms of $E(\kappa)$ by the physical hypothesis, and if $B(\kappa)$ is known, we shall have a single, determinate equation for $E(\kappa)$.

The hypothesis advanced by von Weizsäcker (1948) and Heisenberg (1948 a) is based on the notion that the Fourier components with wave numbers greater than κ act on those with wave numbers less than κ in the manner of molecular friction, that is, as a small-scale rapidly adjusting motion superimposed on a large-scale slowly adjusting motion. The rate of loss of energy of the components with wave numbers between 0 and κ due to a genuine molecular friction is given by the last term in (8.1), so that the postulated expression for $S(\kappa)$ has the form

$$S(\kappa) = 2\mathrm{N}(\kappa)\int_0^\kappa \kappa'^2 E(\kappa')\,d\kappa', \qquad \ldots\ldots(8.2)$$

where $\nu(\kappa)$ is an equivalent viscosity representing the effect of the components with wave numbers larger than κ. The simplest hypothesis about $\mathrm{N}(\kappa)$ is that it has the form of an integral over all wave-number magnitudes greater than κ, and if we assume the integrand to depend only on the wave-number magnitude and the spectrum function (an assumption of expediency, but a natural one) we find

$$\mathrm{N}(\kappa) = c\int_\kappa^\infty [E(\kappa'')/\kappa''^3]^{1/2}\,d\kappa'' \qquad \ldots\ldots(8.3)$$

on dimensional grounds, where c is a constant.

Equation (8.1) now contains E alone, as a function of κ and t, and can be solved when $B(\kappa)$ is known. For values of κ so large that $\int_0^\kappa E(\kappa')\,d\kappa'$ approximates to the total energy, and $B(\kappa)$ approximates to the total rate of supply of energy, we can write the left side of (8.1) as the rate of dissipation ϵ and the corresponding solution is

$$E(\kappa) = \left(\frac{8\epsilon}{9c}\right)^{2/3}\kappa^{-5/3}\left[1 + \frac{8\nu^3}{3c^2\epsilon}\kappa^4\right]^{-4/3}. \qquad \ldots\ldots(8.4)$$

As expected from the implicit use of the ideas of the universal equilibrium explained in § 4, we have recovered the relation $E(\kappa) \propto \epsilon^{2/3}\kappa^{-5/3}$ (which follows from dimensional requirements alone) for values of κ which are large but not comparable with $(\epsilon/\nu^3)^{1/4}$. The expression (8.4) has the virtue of representing a rapid decrease of $E(\kappa)$ (with increasing κ) in the neighbourhood of $\kappa = (\epsilon/\nu^3)^{1/4}$, which is the region of the spectrum in which viscous forces become strong. However, it is not in quantitative agreement with measurements of $E(\kappa)$ at these large wave numbers.

For values of κ large compared with $(\epsilon/\nu^3)^{1/4}$ the relation (8.4) approximates to $E(\kappa) \propto \kappa^{-7}$, whereas measurements (e.g. those reproduced in fig. 3) suggest that $(d \log E/d \log \kappa)$ falls off indefinitely (Stewart and Townsend 1951). The failure of the hypothesis about the energy transfer at these large wave numbers is not surprising, since it specifically assumes the transfer to be from a large-scale *slow* motion to a small-scale *quick* motion, as for frictional processes in general; if we measure the characteristic time of the Fourier components contributing to $N(\kappa)$ by $\left[\int_{\kappa}^{\infty} \kappa'^2 E(\kappa')\, d\kappa'\right]^{-1/2}$ we find that this time decreases, as required, when $E(\kappa) \propto \kappa^{-5/3}$, but increases indefinitely with further increase of κ, contrary to the assumed condition. Townsend (1951 c) has recently proposed a different model of the motion associated with wave numbers large compared with $(\epsilon/\nu^3)^{1/4}$ which seems to account for the observed spectrum more satisfactorily.

The expression (8.4) is a solution of equation (8.1) (with (8.2) and (8.3)) only for values of κ large enough to lie within the equilibrium range. At smaller values of κ the solution for $E(\kappa)$ depends explicitly on both κ and t, as well as on $B(\kappa)$. Heisenberg (1948 b) has suggested applying the above expression for the energy transfer to the case of decaying homogeneous turbulence (for which $B(\kappa) \equiv 0$), and has described approximate analytic forms of solutions which retain their shape during the decay, and which make the total energy decay as t^{-1} as observed. Accurate solutions of the same type have been computed by numerical integration for several Reynolds numbers by Chandrasekhar (1949), and by Proudman (1951), who has also made a comparison of these solutions with the relevant measurements. The comparison is not wholly straightforward in view of the presence of the unknown constant c in the expression (8.3), but it appears that when c is chosen as about 0·4 the theoretical solutions represent the observation well over the range of wave numbers in which most of the energy lies. A more extensive comparison with experiment is needed, but tentatively we can regard (8.2) and (8.3) as providing a good representation of the transfer of energy across the spectrum except at very large and (for kinematical reasons) at very small wave numbers.

(iii) *The Probability Distribution of the Velocity Fluctuations*

Many of the first measurements of turbulent motion were directed towards obtaining the probability distribution (p.d.) of the velocity fluctuation at one point in the fluid. More recently measurements have concerned the joint probability distribution (j.p.d.) of the velocity fluctuations at two fixed points with specified separation. In order to obtain a mathematically complete statistical description of the turbulence we should have to continue this sequence, and obtain the j.p.d. of the velocity at any n points in the fluid; however, incomplete as the information provided by the cases $n = 1$ and 2 may be, it is of considerable general interest. The p.d. of the velocity fluctuations is of course strongly influenced by any lack of spatial uniformity of the turbulence which may exist, and as yet few general conclusions about the details of this influence have been obtained. The following remarks apply to the case of decaying homogeneous turbulence, although where quantities depending on the Fourier components of wave numbers in the equilibrium range alone are concerned, they will have more general significance.

The many measurements of the p.d. of the velocity fluctuation at one point that have been made leave no doubt that it is closely normal, or Gaussian, and little more need be said about this distribution. An error-law distribution is not surprising, since the velocity at one point is the resultant of the influences of a large number of eddies or flow patterns, more or less independent of each other, in the neighbourhood of the point.

There are quite definite departures from normality of the j.p.d. of the simultaneous velocity fluctuations at two points. This is revealed by measurements of, for example, the p.d. of the difference of the velocities at these two points (which would be normal if the j.p.d. were normal). Fig. 11 shows some measurements, made by Townsend, of the probability density function of the space derivative $\partial v_1/\partial x_1$ (which is identical with the p.d. of the difference of the velocities at two neighbouring points), where x_1 and v_1 denote parallel components of the position and velocity vectors. More sensitive indications of the lack of normality are provided by observations of $\overline{(\partial v_1/\partial x_1)^4}/[\overline{(\partial v_1/\partial x_1)^2}]^2$, found to be about 3·6, and of $\overline{(\partial v_1/\partial x_1)^3}/[\overline{(\partial v_1/\partial x_1)^2}]^{3/2}$, found to be about $-0{\cdot}3$ at large

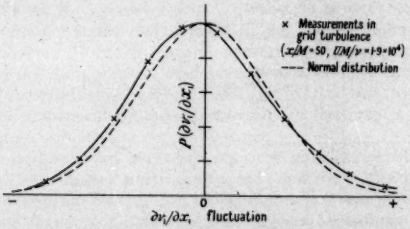


Fig. 11. Probability density function of a velocity derivative.

Reynolds numbers (as compared with the values 3·0 and 0 respectively for a normal distribution). Measurements of similar parameters of the p.d. of $v_1(x_1+r, x_2, x_3)-v_1(x_1, x_2, x_3)$ have been made for the whole range of values of r. When r is large the velocities at the two points are statistically unrelated and we recover the normality of the distribution at a single point, while when r is small we recover the results, already described, for the derivative. Between these two extreme values of r the parameters of the p.d. of the velocity difference vary monotonically with r (Stewart 1951).

The interesting point is that the departures from normality of the p.d. of $v_1(x_1+r, x_2, x_3)-v_1(x_1, x_2, x_3)$ decrease fairly rapidly as r increases, and are quite small when r is still small enough for an appreciable correlation between $v_1(x_1+r, x_2, x_3)$ and $v_1(x_1, x_2, x_3)$ to exist. This can be interpreted (Batchelor 1951) as meaning that the Fourier coefficients associated with wave numbers whose magnitudes are not too large are statistically independent of each other, and that the range of wave-number magnitudes concerned contains an appreciable proportion of the total energy. A consequence of the observation which promises to be useful is that if we are confronted with a complicated statistical quantity involving the velocities at two points, which is known to be determined by the Fourier coefficients associated with small (roughly speaking, the energy-containing range) wave-number magnitudes, we can assume it is related to simpler statistical

quantities in the same way as for a normal j.p.d. of the two velocities. An approximation of this kind has been used to obtain an expression for the covariance of the fluctuating pressures at two points in terms of the covariance of the velocities at the same two points (Heisenberg 1948a, Obukhoff 1949b, Batchelor 1951).

The departure from normality of the p.d. of the difference of the velocities at two points close together is understandable in general terms although not from any rigorous analysis. The relation between the velocities at two points must conform, in some measure, to the equations of momentum and of continuity, which will restrict the possibility that the p.d. is of the pure chance type. The non-linear inertia terms in particular seem likely to impose constraints on the kinds of flow pattern which exist in the neighbourhood of a point. The closer the two points are together, the more will the p.d. of the velocity difference be governed by the requirements of the Navier–Stokes equation, which is consistent with the observation that the departures from normality are greatest for $r=0$. In higher order derivatives the effect of the non-linear inertia forces might be

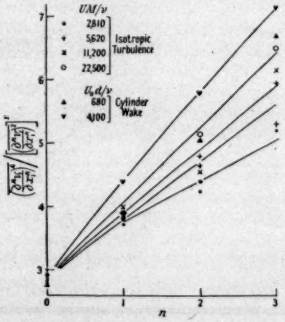


Fig. 12. Flatness factors of velocity derivatives.

expected to be even more accentuated, and it is indeed observed that the p.d. of $\partial^n v_1/\partial x_1^n$ becomes more markedly non-normal as n increases. Fig. 12 shows measurements of the flatness factor,

$$\overline{\left(\frac{\partial^n v_1}{\partial x_1^n}\right)^4} \Big/ \left[\overline{\left(\frac{\partial^n v_1}{\partial x_1^n}\right)^2}\right]^2,$$

for $n=0$, 1, 2 and 3, in the turbulence generated by a square-mesh grid and in a wake (Batchelor and Townsend 1949). For $n=3$ the flatness factor is so high that the p.d. must be very far from being normal. An interpretation of these measurements is that the energy associated with very large wave numbers (and thus contributing to high-order derivatives) is very unevenly distributed in space, and is confined to fairly sharply defined isolated 'activated' regions, somewhat in the manner that, in many flow problems, vorticity is observed to be concentrated in strong isolated vortices. A full discussion of this interesting spottiness of the spatial distribution of fluctuations in high order derivatives of the velocity is beyond our scope here; it must be enough to report that it seems to exist.

(iv) *Generation of Noise by Turbulent Motion*

Although at the time of writing no work on this subject has been published, there are several developments in progress which may be reported briefly. Almost all the existing theoretical research on turbulence assumes that the fluid is incompressible; the analysis is otherwise very complicated, and in any case there are not many physical problems in which the velocity fluctuations are comparable with the local speed of sound. However, with the advent of jet propulsion and the use of very high speed jets on aircraft, and in view of the current interest * in turbulent motion of interstellar gas at Mach numbers of the order of unity, it is now of interest to enquire about the effect of the compressibility of the medium. The equation $\nabla . \mathbf{v} = 0$, or $\mathbf{x} . \mathbf{A}(\mathbf{x}) = 0$, is no longer valid, and is replaced by an equation describing the temporal rate of change of density, together with an equation of state relating pressure and density. The Fourier coefficient $\mathbf{A}(\mathbf{x})$ is no longer required to be perpendicular to the wave number vector, and the resolution of the instantaneous velocity distribution will require longitudinal waves in addition to the transverse waves used for incompressible fluid. These longitudinal waves will tend to propagate as sound waves, although the components of $\mathbf{A}(\mathbf{x})$ transverse and parallel to $\mathbf{x}$ will interact, through the agency of the non-linear inertia terms in the Navier–Stokes equation, and the two components do not in general present separate dynamical problems. We can regard the component of $\mathbf{A}(\mathbf{x})$ parallel to $\mathbf{x}$ as representing the 'noise' component of the turbulence. Moyal (1952) has worked out the various formal kinematical and dynamical relations appropriate to this case in which random noise and random vorticity coexist and interact.

The dynamical problem of determining the amounts of turbulent energy associated with the 'noise' and with the eddy motion is clearly very difficult when these amounts are comparable, or, indeed, in any case in which there is an appreciable flow of energy in both directions between the two modes. However, much of the difficulty disappears if the noise does not react back on the eddy motion and if the Mach number of the eddy motion is low enough for it to be regarded as the same as for an incompressible fluid. These conditions are realized in the case of a turbulent motion, at an appropriately low Mach number, in a finite region of fluid, the boundaries of which permit the noise generated by the turbulence to escape to infinity. The study of problems of this kind has been initiated by Lighthill (1952). He has carried through the analysis for the important case of a turbulent jet discharging into still air, and Proudman (1952) has considered the case of a finite region of fluid in isotropic turbulent motion radiating noise to surrounding stationary fluid.

It follows readily from the equations of continuity and momentum for a compressible fluid that

$$\frac{\partial^2 \rho}{\partial t^2} = \nabla^2 p + \frac{\partial^2 \rho v_i v_j}{\partial x_i \partial x_j} - \frac{4}{3} \mu \nabla . \mathbf{v} \qquad \ldots\ldots (8.5)$$

(where repeated suffixes are summed). The essential step in Lighthill's analysis is to rewrite this equation for the density fluctuations in the form

$$\frac{\partial^2 \rho}{\partial t^2} - a_0^2 \nabla \rho = \frac{\partial F_i(\mathbf{x}, t)}{\partial x_i}, \qquad \ldots\ldots (8.6)$$

* See *Proceedings of the Symposium on the Motion of Gaseous Masses of Cosmical Dimensions*, organized by the International Unions of Astronomy and Theoretical and Applied Mechanics, Paris, 1949; published by Central Aix Documents Office, U.S.A.

where a_0 is the velocity of sound in the surrounding fluid (which is supposed uniform and at rest apart from the motion due to the passage of sound waves through it), and

$$F_i(\mathbf{x}, t) = \frac{\partial}{\partial x_j}[\rho v_i v_j + (p - a_0^2 \rho)\delta_{ij} - \tfrac{4}{3}\nu \nabla \cdot \mathbf{v}\delta_{ij}]. \qquad \ldots\ldots (8.7)$$

In this form the equation describes the propagation of sound waves in a medium in which an extraneous force $\mathbf{F}(\mathbf{x}, t)$ acts on unit volume of the medium at position $\mathbf{x}$ and time t. Outside the region of turbulent motion $\mathbf{F}(\mathbf{x}, t)$ is negligibly small, on the basis of the usual approximations of acoustics, so that the density there satisfies the wave equation. Within the region of turbulent motion $F_i(\mathbf{x}, t)$ is finite principally on account of the contribution from the quadratic term $\partial \rho v_i v_j / \partial x_j$. If $F_i(\mathbf{x}, t)$ is evaluated with the approximation that the turbulent motion is identical with that for a (uniform) incompressible fluid (i.e. neglecting the effects of the noise in this region) we have

$$\frac{\partial F_i(\mathbf{x}, t)}{\partial x_i} = \frac{\partial^2 \rho v_i v_j}{\partial x_i \partial x_j} = -\nabla^2 p, \qquad \ldots\ldots (8.8)$$

showing the dependence of the propagation of the density on the spatial variation of pressure produced by the turbulent motion. We note also the important point that whereas a distributed extraneous force is equivalent in general to a distribution of dipole sources throughout the fluid, the expression (8.7) for $F_i(\mathbf{x}, t)$ is an *exact* divergence (i.e. $\mathbf{F}(\mathbf{x}, t)$ is derivable from a stress system), so that in this case the distributed force is equivalent to a distribution of quadrupole sources.

The particular integral of (8.6) varies with time only if $F_i(\mathbf{x}, t)$ varies with time, and it turns out that the intensity of the sound radiation at positions far from the turbulence depends on both the spatial and temporal distribution of the pressure in the turbulent motion. A knowledge of the statistical properties of these pressure fluctuations is not available, but nevertheless some tentative results about the sound radiation have been obtained with the aid of certain approximations. The acoustic energy flux away from the turbulence, made non-dimensional with the aid of a velocity and a length characteristic of the turbulent motion, is found to vary as the fifth power of the Mach number of the turbulent motion. Lighthill points out that a turbulent motion in which there is a strong mean shearing motion will produce the most noise in general, since the energy flux in such cases varies as the square of the mean rate of shear. The alignments of the distributed quadrupole sources are determined uniquely by the mean shear (where this exists, and generates the turbulence), and the radiation pattern at infinity for a plane shear layer is thus the same as that for a single quadrupole. Proudman's analysis for the case of isotropic turbulence suggests that the Fourier components (of the velocity distribution) belonging to the energy-containing range of wave numbers make an important contribution to the acoustic energy flux. None of these results has yet been tested experimentally.

APPENDIX

ADVANCES IN MEASURING TECHNIQUE

By Dr. A. A. TOWNSEND
Emmanuel College, Cambridge

The great increase in the volume and scope of experimental studies of turbulence over recent years has been accompanied by considerable improvements in methods of measurements, particularly in the technique of the hot-wire anemometer. Comparatively few other measuring instruments have been used, and these mostly for special purposes. This concentration of effort on the technique of the hot-wire anemometer arises from its considerable flexibility, and the property it possesses of delivering an instantaneous record of the turbulent velocity fluctuation. By suitable modifications in construction it has been adapted to measure the various velocity components in turbulent flows of speeds ranging from less than 1 metre sec^{-1} to over 1 000 metre sec^{-1}, and the associated amplifiers and analysing circuits allow the measurement of a wide range of quantities of theoretical interest.

In this Appendix an attempt is made to summarize briefly the principal advances in technique of the last decade but the selection is necessarily a little arbitrary and no claim is made for completeness.

HOT-WIRE CONSTRUCTION

Most of the early workers (Hall 1938, Dryden *et al.* 1937) used hot-wires of platinum, 0·0005–0·0001 in. diameter, usually produced by the Wollaston method, and such wires are still widely used at air velocities which are not too high. Thin platinum wires suffer from lack of mechanical strength, and the tendency, particularly in U.S.A., where supersonic investigations are now very popular, is to use tungsten wires of 0·0002 in. diameter (Kovásznay 1950, Schubauer and Skramstad 1947, van der Hegge Zijnen 1951). Tungsten has the advantage that it is strong and of higher resistivity than platinum, but it has been necessary to develop techniques for welding or plating the wire ends in order to obtain good electrical connections. For measurements of atmospheric turbulence, nickel and platinum wires of diameter around 0·0005 in. (Shiotani and Yamamoto 1950, Richardson 1950), have been used successfully. Both tungsten and nickel wires may be cleaned by washing and brushing, which is not always possible with the thinner platinum wires.

For measurement of fluctuations in the velocity component parallel to the mean stream, a short straight length of wire placed at right angles to the wind direction is always used. The only advance in recent years is that investigators no longer use excessively long wires in an attempt to increase the output of the hot-wire.

For measurement of the cross-stream component, wire assemblies in the form of a cross have been widely used (Liepmann and Laufer 1947, Townsend 1949 a, Corrsin and Uberoi 1949). This form is easily adapted to the simultaneous measurement of the down-stream and one cross-stream component, which is a feature useful for measurement of shear stress and one-point triple correlations in shear flow (Townsend 1949 a).

A further development of the hot-wire anemometer is its application to the direct measurement of turbulent heat transport (Corrsin 1947, Townsend 1949 b). This is possible since a hot-wire responds both to the temperature and the velocity fluctuations, and the relative sensitivities depend on the heating current. If the mean transport is stationary in time, successive measurements at two or three values of the heating current allow computation of the heat flux and the temperature fluctuations. Measurements of heat flux in the wake of a heated cylinder have been found to be consistent with computations from the mean temperature profiles (Townsend 1949 b). In some circumstances successive measurements are not feasible, e.g. in meteorology or in single spot diffusion trials, and then two closely spaced parallel wires operating at different heating currents can be used to provide two electrical outputs, one proportional to the temperature and one to the velocity fluctuation (Townsend 1951 d).

SHEAR STRESS MEASUREMENT

The measurement of the turbulent shear stress by crossed hot-wires has the disadvantage of requiring several successive measurements, and several specialized instruments have been used to simplify the measurement. The most direct is the floating plate used by Sheppard (1947), which measures the tangential stress on the earth's surface by isolating an area and directly measuring the tangential drag on it. The nature of this instrument makes it unsuitable for use on the laboratory scale, but workers in Göttingen have used a small plate which is thermally insulated and electrically heated to measure the shear stress at a wall. This uses the known result that the loss of heat is proportional to the square root of the local velocity gradient, and the instrument may be calibrated in channel flow under a pressure gradient. Satisfactory results are reported with this technique.

CORONA DISCHARGE ANEMOMETER

Fucks (1950) has used the corona discharge between a small sphere and a ring to measure fluctuations in turbulent velocities, and finds that the frequency response is very good, in contrast to earlier instruments using a glow discharge. The sensitive volume exceeds that of the hot-wire anemometer, but the instrument has advantages when rapid response is essential.

HEAT-SPOT TECHNIQUE FOR DIFFUSION MEASUREMENTS

By passing a current pulse of short duration through a short length of thin wire it is possible to introduce into a flow a finite quantity of heat within the time taken for the wire to cool to the ambient temperature. Using wire of 0·0004 cm diameter, the time constant is approximately 0·0003 sec., and the heating is effectively instantaneous. Having produced such a spot, two types of diffusion experiment are possible: the movement of such a spot relative to a similar spot, and the distortion and diffusion of individual spots may be studied. In both experiments suitable statistical methods of measurements must be sought, for it is not practicable to trace completely the motion of a spot. Since means over large numbers of trials are required, measurements of detection probabilities at fixed points over a plane down-stream of the point of origin will supply the required information. Measurements of the probability distributions of spot peak temperatures in turbulent flows have been made using this technique (Townsend 1951 d).

ATMOSPHERIC TURBULENCE

Turbulence measurement in the atmosphere poses special problems, some due to the uniqueness and unrepeatability of any set of observations, some to the extremely large turbulent intensities, and some to the long time scale of the fluctuations. The use of wire assemblies for measurement of velocity and temperature fluctuations is now common, but methods of analysis are many and varied. One interesting development is the study of atmospheric turbulence by measuring the variability of sound or light transmission through the atmosphere (Richardson 1950). Preliminary investigations seem to show that this may be a convenient way of studying atmospheric turbulence of a large scale.

ELECTRICAL EQUIPMENT

Amplifiers. The wide range of measurements now attempted with hot-wire anemometers has led to considerable improvement in the performance and convenience of signal amplifiers. A good amplifier should have a stable gain factor and a low inherent noise level, incorporate a circuit compensating the response for the loss due to thermal lag of the hot-wire, develop sufficient undistorted output to operate a vacuum thermojunction operating a millivoltmeter, and preferably be completely operated from the main electricity supply. Several designs are in use fulfilling these requirements. For stability and good low-frequency response it is usual to use push-pull amplification throughout, a procedure that relaxes considerably the necessity for low internal resistance, stability and freedom from hum of the power supply. An important feature is the use of low-pass filters to cut off abruptly the frequency response at a predetermined value. If the filter is omitted, a great deal of noise appears in the amplifier output, mostly at very high frequency, and no loss of turbulent signal need arise from cutting off all frequencies above a limit, the choice of which naturally depends on the individual turbulent flow. For most laboratory flows at air speeds near $1\,000\,\mathrm{cm\,sec^{-1}}$ no appreciable loss results from a cut-off at 5 000 c/s. Measurements at supersonic speeds require responses to over 10^5 c/s, and the amplifier and filter must be designed accordingly. Such extended response raises the noise level very appreciably, but few workers have taken the obvious step of including an input transformer to improve the signal–noise ratio. The only reported measurements using an input transformer are by Schuh and Winter (1948), who used it to measure the background turbulence in a low-turbulence wind tunnel.

If the turbulent fluctuations exceed 10% of the mean wind there is appreciable non-linearity of the hot-wire response, and either this must be eliminated or corrections for it must be made. Making corrections requires a knowledge of the statistical distribution of the fluctuations, and it is simpler to introduce non-linear distortion of amount sufficient to straighten the wire response. A practical circuit for use in atmospheric turbulence based on a circuit with a square-law response incorporated in a direct-coupled amplifier has been used at Cambridge by Ellison. A similar square-law circuit is used to measure instantaneous and mean squares of the velocity.

Constant resistance operation. While most workers have been content to use the conventional system of compensating for the hot-wire thermal lag after it has occurred, several attempts have been made to provide a feedback circuit which tends to produce a constant wire resistance, i.e. constant wire temperature,

independent of wind speed (Ossofsky 1948, Weske 1948, Tournier, Laurenceau, Heubès and Seigneurin 1951). If the adjustments of the circuit can be carried out very rapidly, no temperature changes occur and no compensating circuits are necessary. Early circuits used direct coupled amplifiers to amplify the out-of-balance voltage from the hot-wire bridge, and the amplified voltage was added to the bridge supply in such a direction that increased temperature led to reduced heating current. Self-oscillation prevented reliable operation on this plan, but more recently several workers have used the carrier-frequency version of this plan successfully. The advantages of the scheme are most apparent when the normal methods of determining the compensation setting fail or when the fluctuations are so large that the thermal response time varies appreciably. Applications to atmospheric turbulence and to turbulent flow near walls seem to be the most promising.

Intensity measurement. It is usual to measure intensities with a vacuo-junction milliammeter combination, but measurements of atmospheric turbulence often require means over time intervals longer than the time constant of the thermojunction. Richardson (1950) has attempted to remedy this by rectifying the fluctuations and arranging an electronic integrator to operate a mechanical counter. The counting rate is a measure of the intensity, but is rather dependent on the statistical distribution of the velocity fluctuations. A better system now under trial, uses periodical sampling of short-period averaging circuits, with condenser storage and subsequent measurement with an electronic voltmeter.

Spectral analysis of the output from a single hot-wire. Most of this analysis depends on the validity of the approximation introduced by G. I. Taylor that the turbulent velocity pattern varies so slowly that the velocity fluctuations recorded by a fixed hot-wire are due entirely to motion of this pattern past the wire, and only to a negligible extent to change in the pattern. The approximation is good in free turbulence if the level is low but fails at high levels of turbulence, particularly if there are large pressure gradients, as, for example, in flow near a wall. In free turbulence the identity of the auto-correlation function of the hot-wire signal and the longitudinal spatial correlation has been confirmed by several workers (Taylor 1938b, Stewart and Townsend 1951).

Measurement of the dissipation parameter λ. The characteristic length λ, introduced by G. I. Taylor, is defined by $1/\lambda^2 = -f_0'' = \overline{(\partial u/\partial x)^2}/\overline{u^2}$. To the usual approximation $\partial u/\partial x = -U^{-1}\partial u/\partial \tau$, where τ is experimental time, so that λ^2 may be computed from the ratio of the mean square of u to the mean square of its time derivative. Simple circuits exist for the purpose of differentiating electrical signals with respect to time, and the measurement is simple and accurate (Townsend 1947a). If the turbulent motion is isotropic, measurements of $\overline{u^2}$ and λ^2 determine the local viscous dissipation of energy ϵ, since $\epsilon = 15\nu\overline{u^2}/\lambda^2$. There is considerable evidence that the local energy dissipation is given with sufficient accuracy by the same expression in shear flow, as suggested by the universal equilibrium theory, and use of this or similar expressions is now common in experimental studies of shear flow (Townsend 1949a, Laufer 1950b).

An alternative method of determining λ which gives almost identical results, though its theoretical backing is less secure (Laufer 1950b), is to measure the frequency n_0 with which the trace of u crosses the zero fluctuation position. Circuits which produce a pulse every time this happens can be devised, and the pulse rate can then be measured with an electronic counter. If it can be assumed

that u and $\partial u/\partial\tau$ are statistically independent, then $\lambda = U/2\pi n_0$. The assumption of statistical independence is a theoretical difficulty, but the method has the advantage that estimates of λ can be obtained rapidly from photographic records, in particular meteorological and flight records. In laboratory work the advantages are less obvious.

Measurement of scale. The longitudinal integral scale $L = \int_0^\infty f(r)\,dr$ is a measure of the size of the larger eddies of turbulence. It is related to the one-dimensional intensity spectrum functions of u by

$$L = \tfrac{1}{2}\pi\phi(0)/\overline{u^2} = \tfrac{1}{2}\pi U\psi(0)/\overline{u^2}$$

(ϕ and ψ are spectrum functions in x and τ). The scale is then determined by the spectral intensity at zero wave number or zero frequency. If a low-pass filter is inserted in the amplifier, then from the insertion loss the scale may be calculated (Townsend 1947 a).

Successive differentiations. In principle, the wire output may be differentiated with respect to time any given number of times, but the noise level rapidly becomes troublesome and, even if filters are used to restrict the frequency range to be no more than the minimum necessary to include all the turbulence spectrum, three successive differentiations are the limit without using an input transformer (Batchelor and Townsend 1949). Use of an input transformer might permit four successive differentiations.

Measurement of spectra. Considerable interest is attached to the form of the spectrum function, particularly at large wave numbers, and accurate measurements are very useful. A considerable difficulty in direct measurement is the handling of the whole intensity by the spectral analyser without overloading when the energy of the high-frequency components being measured is extremely small. It is more convenient to measure the spectra of the successive derivatives of the signal, for then the range of measured spectral intensities and the possibility of overload is much reduced (Stewart and Townsend 1951). Alternatively, suitable filters may be used to eliminate unwanted frequencies.

At low frequencies the statistical fluctuations in spectral intensity become so great that it is better to measure either spatial correlations or auto-correlations and to compute the spectrum function from these

Triple and quadruple velocity products. The measurements of the mean value of a triple or quadruple velocity product, e.g. $\overline{u_1^2 u_2}$ or $\overline{u_1^2 u_2^2}$, can usually be reduced to two operations: (*a*) the production of a signal proportional to the instantaneous square of the input voltage, (*b*) the measurement by a vacuum thermojunction of the mean square of the sum (or difference) of two signals. The second operation is a standard one, but the first operation requires the use of non-linear circuit elements. The curvature of triode characteristics near plate current cut-off may be used to produce the required characteristic very simply and effectively, and suitable connection of pentodes or hexodes will also serve the purpose (Townsend 1947 a). For measurements at low values of the product it is advantageous to use a double circuit which subtracts the relevant intensities continuously, and this reduces considerably the statistical fluctuations (Stewart 1951).

These circuits have been useful in the study of turbulent energy diffusion in shear flow as well as in isotropic turbulence (Townsend 1949 a, b, 1951 a).

Intermittency. The intermittent nature of turbulence in wake flow has been studied by constructing a signal whose value is zero in the absence of turbulence and a constant value when the hot-wire is in a region of fully developed turbulence. Using this signal it is possible to measure directly the intermittency factor, and to examine the size and nature of the fully turbulent regions (Townsend 1949 b).

Magnetic tape recording and delay. The availability of high performance magnetic tape recorders has made possible the recording and storage of hot-wire signals either directly or as a modulated carrier. Possibly the most promising application is to record atmospheric turbulence fluctuations and later to analyse the record, perhaps running the tape through at a greatly increased speed. In this way a whole spectrum may be measured using the same run, which is an important feature when local conditions are variable, as they usually are in the atmosphere. Moreover, an additional measurement may be made later, even if, at the time of recording, this particular type of measurement was not intended.

The possibility of recording has not the same advantages for laboratory turbulence, but the use of it to produce a time delay allows measurements of correlations in time, particularly the auto-correlation of the velocity at a point moving with the mean stream (Favre *et al.* 1950).

ACKNOWLEDGMENTS

Figures 1, 9 and 10 are reproduced by permission of the Institute of Aeronautical Sciences, and figs. 2, 4, 5, 7 and 12 by permission of the Royal Society.

REFERENCES

Batchelor, G. K., 1947, *Proc. Camb. Phil. Soc.*, **43**, 533; 1948, *Proc. 7th Intern. Cong. Appl. Mechs.* (London), Intro. Vol.; 1949 a, *Proc. Roy. Soc.* A, **195**, 513; 1949 b, *Aust. J. Sci. Res.* A, **2**, 437; 1950 a, *Quart. J. Roy. Met. Soc.*, **76**, 133; 1950 b, *Proc. Roy. Soc.* A, **201**, 405; 1951, *Proc. Camb. Phil. Soc.*, **47**, 359; 1952, *Proc. Roy. Soc.*, in the press.

Batchelor, G. K., and Stewart, R. W., 1950, *Quart. J. Mech. Appl. Math.*, **3**, 1.

Batchelor, G. K., and Townsend, A. A., 1947, *Proc. Roy. Soc.* A, **190**, 534; 1948 a, *Ibid.*, **193**, 539; 1948 b, *Ibid.*, **194**, 527; 1949, *Ibid.*, **199**, 238.

Burgers, J. M., 1948, *Advances in Applied Mechanics*, Vol. 1 (New York: Academic Press), p. 171.

Chandrasekhar, S., 1949, *Proc. Roy. Soc.* A, **200**, 20.

Corrsin, S., 1947, *Rev. Sci. Instrum.*, **18**, 469; 1949, *J. Aero. Sci.*, **16**, 757; 1950, *Ibid.*, **17**, 417.

Corrsin, S., and Uberoi, M. S., 1949, *N.A.C.A. Tech. Note*, No. 1865; 1950, *Ibid.*, No. 2124.

Dryden, H. L., 1948, *Advances in Applied Mechanics*, Vol. 1 (New York: Academic Press), 8.

Dryden, H. L., Schubauer, G. B., Mock, W. C., and Skramstad, H. K., 1937, *N.A.C.A. Tech. Rep.*, No. 581.

Fage, A., 1936, *Rep. Prog. Phys.*, **3**, 26 (London: Physical Society).

Favre, A., Gaviglio, J., and Dumas, R., 1950, *Note Tech. O.N.E.R.A.*, No. 13/522A.

Fucks, W. A., 1950, *Z. Naturforsch.*, **5 a**, 89.

Goldstein, S., 1951, *Proc. Camb. Phil. Soc.*, **47**, 554.

Hall, A. A., 1938, *Rep. & Memo. Aero Res. Comm.*, No. 1842.

van der Hegge Zijnen, B. G., 1951, *Appl. Sci. Res.* A, **2**, 351.

Heisenberg, W., 1948 a, *Z. Phys.*, **124**, 628; 1948 b, *Proc. Roy. Soc.* A, **195**, 402.

Hinze, J. O., and van der Hegge Zijnen, B. G., 1949, *Appl. Sci. Res.* A, **1**, 435.

Inoue, E., 1950, *Geophys. Note, Tokyo Univ.*, **3**, No. 34.

von Kármán, T., and Howarth, L., 1938, *Proc. Roy. Soc.* A, **164**, 192.

von Kármán, T., and Lin, C. C., 1951, *Advances in Applied Mechanics* (New York: Academic Press), Vol. 2.

KHINTCHINE, A., 1933, *Math. Ann.*, **109**, 604.
KOLMOGOROFF, A. N., 1941 a, *C. R. Acad. Sci. U.R.S.S.*, **30**, 301; 1941 b, *Ibid.*, **32**, 16; 1949, *Dokl. Akad. Nauk, S.S.S.R.*, **66**, 825.
KOVÁSZNAY, L. S. G., 1950, *J. Aero. Sci.*, **17**, 565.
LAUFER, J., 1950 a, *J. Aero. Sci.*, **17**, 277; 1950 b, *N.A.C.A. Tech. Note*, No. 2123.
LIEPMANN, H. W., and LAUFER, J., 1947, *N.A.C.A. Tech. Note*, No. 1257.
LIGHTHILL, J., 1952, *Proc. Roy. Soc.* A, in the press.
MEGAW, E. C. S., 1950, *Nature, Lond.*, **166**, 1100.
MOYAL, J. E., 1949, *J. Roy. Statist. Soc.*, Ser. B, **11**, 150; 1952, *Proc. Camb. Phil. Soc.*, in the press.
OBUKHOFF, A. M., 1941, *C. R. Acad. Sci., U.R.S.S.*, **32**, 19; 1949 a, *Izv. Akad. Nauk, S.S.S.R.*, Ser. Geogr. Geofiz., **13**, 58; 1949 b, *Dokl. Akad. Nauk, S.S.S.R.*, **66**, 17.
ONSAGER, L., 1945, *Phys. Rev.*, **68**, 286.
OSSOFSKY, E., 1948, *Rev. Sci. Instrum.*, **19**, 881.
PROUDMAN, I., 1951, *Proc. Camb. Phil. Soc.*, **47**, 158; 1952, *Ibid.*, in the press.
RICHARDSON, E. G., 1950, *Proc. Roy. Soc.* A, **203**, 149.
ROBERTSON, H. P., 1940, *Proc. Camb. Phil. Soc.*, **36**, 209.
SCHUBAUER, G. B., and KLEBANOFF, P. S., 1950, *N.A.C.A. Tech. Note*, No. 2133.
SCHUBAUER, G. B., and SKRAMSTAD, H. K., 1947, *J. Res. Nat. Bur. Stand., Wash.*, **38**, 251.
SCHUH, H., and WINTER, K. G., 1948, *Proc. 7th Intern. Cong. Appl. Mechs.* (London), Vol. 2, 617.
SHEPPARD, P. A., 1947, *Proc. Roy. Soc.* A, **188**, 208.
SHIOTANI, M., and YAMAMOTO, G., 1950, *Geophys. Mag.*, **21**, 134.
SQUIRE, H. B., 1948, *Phil. Mag.*, **39**, 1.
STEWART, R. W., 1951, *Proc. Camb. Phil. Soc.*, **47**, 146.
STEWART, R. W., and TOWNSEND, A. A., 1951, *Phil. Trans. Roy. Soc.* A, **243**, 359.
TAYLOR, G. I., 1935 a, *Proc. Roy. Soc.* A, **151**, 421; 1935 b, *Z. Angew. Math. u. Mech.*, **15**, 91; 1938 a, *Proc. Roy. Soc.* A, **164**, 15; 1938 b, *Ibid.*, **164**, 476.
TAYLOR, G. I., and BATCHELOR, G. K., 1949, *Quart. J. Mech. Appl. Math.*, **2**, 1.
TOURNIER, M., LAURENCEAU, P., HEUBÈS, J. J., and SEIGNEURIN, A., 1951, *Rech. Aéronaut.*, No. 19, 11.
TOWNSEND, A. A., 1947 a, *Proc. Camb. Phil. Soc.*, **43**, 560; 1947 b, *Proc. Roy. Soc.* A, 190, 551; 1948, *Aust. J. Sci. Res.* A, **1**, 161; 1949 a, *Proc. Roy. Soc.* A, **197**, 124; 1949 b, *Aust. J. Sci. Res.* A, **2**, 451; 1950, *Phil. Mag.*, **41**, 890; 1951 a, *Proc. Camb. Phil. Soc.*, **47**, 375; 1951 b, *Quart. J. Mech. Appl. Math.*, **4**, 38; 1951 c, *Proc. Roy. Soc.* A, **208**, 534; 1951 d, *Ibid.*, in the press.
TRUBRIDGE, G. F. P., 1934, *Rep. Prog. Phys.*, **1**, 43 (London: Physical Society).
VON WEIZSÄCKER, C. F., 1948, *Z. Phys.*, **124**, 614.
WESKE, J. R., 1948, *N.A.C.A. Tech. Note*. No. 880.
WIENER, N., 1930, *Acta Math.*, **55**, 117.
YAGLOM, A. M., 1949, *Dokl. Akad. Nauk, S.S.S.R.*, **69**, 743.

A full account of the mathematical and physical developments that are related, directly or indirectly, to homogeneous turbulence (which includes most of th subject matter of §§ 1–6 and 8 of this report) is given in a monograph to appear soon : ***The Theory of Homogeneous Turbulence***, by G. K. Batchelor (Cambridge: University Press)

FERRITES

By A. FAIRWEATHER, F. F. ROBERTS and A. J. E. WELCH *
Post Office Research Station, Dollis Hill, London

CONTENTS

		PAGE
§ 1.	Introduction	142
§ 2.	Historical	143
§ 3.	The structure of ferrites	143
	(i) The spinel crystal lattice	143
	(ii) ' Normal ' and ' inversed ' spinels	144
	(iii) Solid solution effects in ferrites	146
	(iv) ' Gamma ' ferric oxide	147
	(v) The preparation of ferrites	147
§ 4.	Saturation magnetization and static paramagnetic susceptibility	148
§ 5.	Nature of the magnetic interactions	153
§ 6.	Permeability–frequency behaviour	156
§ 7.	' Damping '	160
§ 8.	Magnetic losses at low frequencies and low levels	161
§ 9.	Dielectric and semiconducting properties	162
	(i) Engineering considerations	162
	(ii) Dielectric properties	164
	(iii) Semiconducting properties	166
§ 10.	Concluding remarks	168
	Appendix : Néel's theory of ferrimagnetism	168
	References	170

ABSTRACT. With d.c. electrical resistivities several orders of magnitude greater than those of the ferromagnetic metals, and magnetic saturation intensities in some cases less than one order of magnitude smaller, the ferrites gave promise of greater ' effective permeability ' and lower ' eddy-current ' loss in high-frequency magnetic fields. Attempts to realize these advantages have had limited success, but have led to a deeper understanding of the ferromagnetic and antiferromagnetic behaviour of non-metallic substances. In the present report, a description of the physical and chemical structure of the ferrites, and of their preparation, is followed by an outline of Néel's theory of ferrimagnetism, and of the Kramers–Anderson theory of superexchange spin coupling. Several possible mechanisms which may account for the observed rather steep fall of permeability with increase of frequency, and for the associated energy absorption, are then discussed. The dielectric properties of ferrites are also of interest in high-frequency applications. The observed very strong dispersion and absorption may be interpreted in terms of the inhomogeneous microphysical structure of the sintered material, which is known to be sensitive to the oxygen content and heat-treatment conditions. The very low-frequency dielectric absorption may be due to adsorbed moisture in the material. The semiconducting properties are briefly discussed. An extensive bibliography is included.

§ 1. INTRODUCTION

THE ferrites first attracted engineering interest as non-metallic ferromagnetic materials for applications at frequencies where eddy currents have undesirable consequences in the usual metallic materials. In the present report, however, the ferrites will be examined, primarily for the light they may throw on the general subject of ferromagnetic and, to a lesser extent, dielectric behaviour. Much of the new information in magnetism, which the study of the ferrites is

* Royal College of Science, South Kensington.

revealing, cannot be brought satisfactorily into the theoretical picture that has been built up for magnetic metals. While this report is not intended to be in any sense a sequel to two earlier very full reports on ferromagnetism (Stoner 1948, 1950), questions related to the frequency variation of magnetic permeability will be discussed in a rather broader setting than is implied by the title ' Ferrites ', and some brief reference to the behaviour of metals under comparable conditions will be made.

§ 2. HISTORICAL

Ferromagnets made of natural magnetite, Fe_3O_4 were known in antiquity, possibly before iron itself was put to any magnetic use. Magnetite, also known as ferroso-ferric oxide, may be written chemically as ferrous ferrite, $Fe^{II}\,Fe_2{}^{III}O_4$ and, as will be seen later, has the same fundamental crystal structure as the ferrites in general, in which the ferrous ion Fe^{2+} is replaced by the ions of certain other bivalent metals. Among the earliest magnetic measurement on magnetites are those of Weiss (1896), who found its saturation magnetization per unit volume comparable with that of nickel, and its Curie temperature between those of nickel and iron. The d.c. electrical resistivity of magnetite is about 1 000 times, and that of other ferrites 10^4–10^{11} times that of iron, so that eddy currents will not be expected to give trouble in transformer cores made of these materials until frequencies are reached much higher than those significant for similar cores of iron. Patents were taken out by Hilpert (1909) for the use of ferrites in such an application. Other early work on ferrites was reported by Wologdine (1909), Hilpert (1909) and Quittner (1909); but apart from chemical work (to be referred to later) and a few structure determinations, there then seems to have been a 20-year period of relative inactivity in this field, presumably due to the satisfaction of engineering requirements by improved sheet steels and alloys, and to the relatively uninspiring state of magnetic theory pending the application of the new quantum mechanics.

The importance of an accurate knowledge of composition and structure was then beginning to be realized (though not adequately appreciated until much later) and more comprehensive investigations were being commenced by Weiss and Forrer (1929), Forestier (1931), Millar (1929), Ching Hsien Li (1932) and especially by Barth and Posnjak (1932).

§ 3. THE STRUCTURE OF FERRITES

(i) *The Spinel Crystal Lattice*

The structures of many metal oxides, particularly of metals giving small polyvalent cations, may be considered broadly as close-packed lattices of oxygen ions with the necessary cations disposed in some regular manner in the interstices. The *spinels* (named after the well-known mineral, spinel, $MgAl_2O_4$) form an extensive group of such oxides having a relatively simple crystal structure.

The composition of an ideal spinel may always be represented by a formula of the type XY_2O_4, in which X and Y denote different metal cations ; X and Y are usually derived from different metal atoms, but in a few cases (e.g. magnetite, $Fe^{II}Fe_2{}^{III}O_4$) only the valencies differ. The total positive charge on one X cation and two Y cations must clearly be eight units; this allows the valency combinations 6–1 (rare), 4–2, and 2–3, the first and second figures representing the valencies (i.e. positive charges) of X and Y cations, respectively. All pure ferrites having the spinel structure are necessarily 2–3 spinels.

The crystal lattice arrangement characteristic of any spinel is shown in perspective in fig. 1 (due to Verwey and Heilmann 1947), which particularly emphasizes the cubic symmetry of the structure; the threefold symmetry axes along the cube diagonals are clearly discernible. The unit cell contains thirty-two oxygen anions (shown as large spheres in fig. 1) and twenty-four cations. These cations have environments of two distinct types, according to the type of interstice in which they are accommodated ; eight of them (shown as black spheres in fig. 1) are surrounded tetrahedrally by four oxygen ions, and the remaining sixteen (shaded spheres) by six oxygen ions placed octahedrally. The presence in the cell of eight tetrahedral and sixteen octahedral interstices in which cations may be accommodated is a requirement of the symmetry of the spinel lattice ; the distribution of the X and Y ions among these interstices is considered in § 3 (ii).

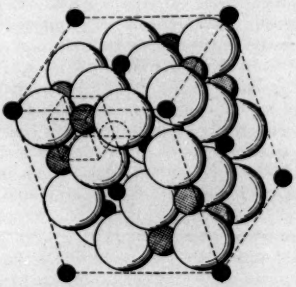


Fig. 1. The spinel crystal structure (*after* Verwey and Heilmann 1947).

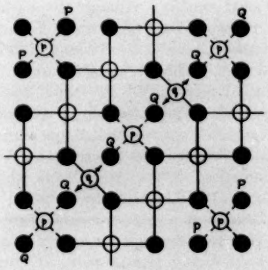


Fig. 2. The spinel lattice (*after* W. L. Bragg).

Another representation of the spinel lattice (fig. 2, due to W. L. Bragg) serves better in showing certain characteristic features of the structure. The figure shows a section through a plane containing oxygen ions and cations in octahedral interstices, parallel to the front face of the cube in fig. 1 and lying a short distance behind it. The black circles represent oxygen ions, and the open and crossed circles denote cations in tetrahedral and octahedral interstices, respectively. For clarity, several oxygen ions lying in planes above (P) and below (Q) the plane of the section, and the cations in tetrahedral interstices slightly above (p) and slightly below (q) this plane, have been inserted in fig. 2. The important feature brought out by fig. 2 is that cations in octahedral interstices form continuous chains running right through the structure, with a relatively small distance between adjacent ions.

(ii) '*Normal*' *and* '*Inversed*' *Spinels*

The spinel structure must contain eight X and sixteen Y cations distributed among eight tetrahedral and sixteen octahedral interstices in the unit cell. It is therefore reasonable to conclude that the X ions will probably occupy tetrahedral and the Y ions octahedral interstices, an arrangement which fully satisfies the requirements of lattice symmetry. It was shown, however, by Barth and Posnjak (1932) that symmetry requirements are also met by another arrangement in which

half the Y ions occupy tetrahedral interstices, the octahedral interstices being filled by the X ions and the other half of the Y ions, distributed at random. Verwey and Heilmann (1947) have termed the latter arrangement ' inversed ', in contrast to the ' normal ' arrangement described first. Verwey and Heilmann have emphasized that the inversed and normal cation arrangements are merely the limits of a continuous range of possible cation distributions satisfying spinel symmetry.

Distinction between the normal and inversed cation arrangements in a given spinel can be made by careful measurements of x-ray diffraction intensities, provided that the X and Y cations have appreciably different scattering power for x-radiation employed. By this method Barth and Posnjak (1932) have shown that magnesium ferrite $MgFe_2O_4$ has the inversed structure. Verwey and Heilmann (1947), who give no details of their intensity measurements, similarly find the inversed structure in cupric ferrite $CuFe_2O_4$, but obtain normal structures in zinc and cadmium ferrites $ZnFe_2O_4$ and $CdFe_2O_4$. Since zinc and cadmium ferrites differ from all other known spinel-type ferrites in being non-ferromagnetic, this observation leads to the conclusion that ferromagnetism in a spinel ferrite is associated with the inversed cation arrangement.

A careful comparison of the values of lattice constants a for a number of spinels has led Verwey and Heilmann (1947) to conclude that an inversed spinel should have a some 0·07 Å less than a normal spinel containing the same cations; this rule affords evidence for an inversed cation distribution in manganous ($MnFe_2O_4$), ferrous ($FeFe_2O_4$ magnetite), cobaltous ($CoFe_2O_4$) and nickel ($NiFe_2O_4$) ferrites, all of which are ferromagnetic. In these four cases x-ray intensity measurements cannot readily be used to distinguish the normal and inversed structures, because the bivalent ions differ little in scattering power from the ferric ion. Lattice constant data have been further discussed by Andrews (1951), but Nicks and Welch (1951) consider that lattice constant data alone do not afford a reliable distinction between the two cation arrangements. The general association of ferromagnetism with the inversed structure has therefore a somewhat insecure experimental foundation ; in particular, small but significant departures from any idealized cation distribution would probably escape detection by the techniques of structural investigation so far applied.

In the inversed spinel structure the chains of octahedral interstices (see § 3 (i)) contain equal numbers of X and Y cations, which in a statistical distribution must occur alternately; this fact may have some bearing on ionic interactions related to ferromagnetic behaviour. In magnetite the octahedral cation chains comprise alternate Fe^{2+} and Fe^{3+} ions, between which electron exchange is readily conceived to occur; this accounts for the electronic conductivity of magnetite (Verwey and de Boer 1936, Verwey, Haayman, and Romeijn 1947, Verwey, de Boer and van Santen 1948). The magnetite lattice is of particular interest because its supposedly inversed cation arrangement could be converted into the normal arrangement merely by shifting eight electrons per unit cell from octahedral to tetrahedral cations.

An unusual compound of lithium and ferric oxides, $LiFe_5O_8$ (Hoffmann 1938, Barblan, Brandenburger and Niggli 1944), is ferromagnetic and has the spinel structure; this compound is distinct from the more widely known ' lithium ferrite ', $LiFeO_2$, which is not ferromagnetic and has a rocksalt crystal lattice. $LiFe_5O_8$ is believed to have an inversed cation arrangement (Gorter 1950 a, b),

with Fe^{3+} ions in tetrahedral positions, and Fe^{3+} and Li^{+} ions in octahedral positions.

The occurrence of normal and inversed cation arrangements in a range of spinels, some of them ferromagnetic, has been discussed briefly by Kordes and Röttig (1951).

(iii) *Solid Solution Effects in Ferrites*

Some solid solubility of the component oxides (XO and Y_2O_3) in 2–3 spinels of stoichiometric composition (XY_2O_4) has been reported frequently. Van Arkel, Verwey and Van Bruggen (1936), basing their conclusions on detection of the undissolved solute phase by x-ray powder photographs, have reported quite wide composition ranges within which a monogeneous spinel phase is thought to exist. Later workers, using the more precise technique of plotting the variation of lattice constant with composition, have found narrower solid-solution ranges; Berger (1948) indicates that the spinel structure is stable in zinc ferrite between 50 and 64·5 mol-% ferric oxide, at 1200° c, and Kingsnorth and Welch (1951) report limits of 42 and 53 mol-% ferric oxide in magnesium ferrite at the same temperature.

The mode of incorporation of small quantities of the component oxides in solid solution in the spinel lattice is not clearly established, but solid solutions containing excess of the tervalent oxide are often regarded as containing some Y ions substituted for an equivalent number of X ions in the ideal spinel structure (Hägg 1935, Hägg and Söderholm 1935). Thus a solid solution of ferric oxide in zinc ions in the stoichiometric ferrite is replaced by two-thirds of that number of ferric ions; the remaining one-third of the zinc-ion sites necessarily remain empty on account of the greater charge of the ferric ions. This mechanism of solid-solution formation is in accord with the observed changes in density and lattice constant, but no direct proof has yet been possible of the detailed changes of cation distribution accompanying it. Nor has a satisfactory account been given of the structure of spinel solid solutions containing excess of the bivalent oxide.

Solid solubility of stoichiometric spinels in other such spinels is usually complete. Substitutional solid solution of a normal 2–3 spinel in a second normal 2–3 spinel (or of one inversed 2–3 spinel in another) is readily visualized; the bi- or ter-valent ions of different species are considered to be statistically distributed over the lattice sites customarily occupied by bi- and ter-valent ions, respectively, in the pure spinels. The existence of a super-lattice (i.e. some special distribution of the different bi- or ter-valent cations among the sites normally available to them) is not excluded, but does not appear to have been detected in any known case.

Substitutional solid solution between a normal and an inversed spinel is less readily conceived, because it must involve some transition between the two characteristic cation arrangements. This transition could be continuous, a progressive change in the arrangement occurring throughout the composition range between the two pure spinels, or it could occur discontinuously over narrower ranges of composition. Gorter (1950 a, b) has found by x-ray measurements that in solid solutions of zinc ferrite (normal structure) in other bivalent metal ferrites having inversed structures, the zinc ions occur in tetrahedral sites, as in pure zinc ferrite. This indicates some discontinuous transition in the cation arrangement. Nicks and Welch (1951) have found similar indications of a discontinuous change in solid solutions of magnesium ferrite ($MgFe_2O_4$, normal). Changes in cation

distribution in such cases merit further study, but are not readily followed because many of the interesting cations have comparable x-ray scattering power and are not easily located by customary x-ray techniques.

(iv) *'Gamma' Ferric Oxide*

'Gamma' ferric oxide (γ-Fe_2O_3), a ferromagnetic form of the oxide obtained by special methods of preparation (see § 3 (v)), has a spinel structure giving an x-ray powder photograph closely resembling that of magnetite Fe_3O_4. The occurrence of the spinel structure in a simple tervalent oxide (also observed in the parallel case of γ-Al_2O_3) involves an abnormally low proportion of cations in the lattice. According to Hägg (1935) and Verwey (1935) the 24 cation sites in the ideal spinel unit cell are occupied, on the average, by $21\frac{1}{3}$ ferric ions, an average of $2\frac{2}{3}$ of the available cation sites in each cell remaining empty. This view implies that the 'gamma' form of ferric oxide could be regarded as the ultimate limit of solid solution of ferric oxide in magnetite, the structural mechanism of solution being that outlined in § 3 (iii). David and Welch (1951) have shown that Hägg's structure is scarcely adequate to account for all the properties of the oxide; the presence of some combined water in the lattice appears to be essential to its stability. Moreover, ferric oxide shows inappreciable solid solubility in magnetite, suggesting that cation defects in the magnetite lattice are not readily introduced by the supposed mechanism. As found by David and Welch (1951) and previous investigators, the powder x-radiogram of 'gamma' ferric oxide contains additional lines which indicate some special distribution of the supposed vacant cation sites.

(v) *The Preparation of Ferrites*

The basic method of preparation of the ferrites comprises thorough mixing of weighed quantities of the pure component oxides, and firing of the mixture at a high temperature, frequently after pressing into suitable compacts. Snoek (1947) and his collaborators have carried out extensive investigations of the effect of conditions of preparation on the magnetic properties of the resulting ferrite. These are particularly affected by the temperature of firing and the oxidizing or reducing character of the furnace atmosphere, and by the mode of cooling.

The preparation of single crystals of ferrites involves special techniques (see, for example, Galt, Matthias and Remeika 1950).

Some dissociation of ferrites with loss of oxygen occurs at high temperatures, even on heating in air or oxygen at normal pressures. This oxygen loss has been found by Snoek (1947) and others to have a considerable effect on the magnetic properties of the ferrites (§ 8). The oxygen loss must be accompanied by some reduction of ferric ions (Fe^{3+}) to ferrous ions (Fe^{2+}), but the proportion of ions so reduced is so small that x-ray techniques do not reveal any structural change in the process. It has been assumed that reduction of iron in a spinel is accompanied by formation of magnetite, $Fe^{II}Fe_2^{III}O_4$ in solid solution in the parent spinel; this assumption is justified by the work of Kingsnorth and Welch (1951).

'Gamma' ferric oxide is prepared by special methods avoiding the use of high temperatures. Perhaps the best is to oxidize precipitated magnetite with air or oxygen at about 250° c. David and Welch (1951) have investigated this process and have shown that magnetite prepared by 'dry' methods cannot be oxidized to the 'gamma' oxide, the presence of some combined water being essential.

§4. SATURATION MAGNETIZATION AND STATIC PARAMAGNETIC SUSCEPTIBILITY

Experiment shows that the low-temperature saturation magnetization of all the ferromagnetic ferrites, of all the ferromagnetic metals except gadolinium, is considerably less than that to be expected if all the transition group atoms in the structure contributed their full magnetic moment, as inferred from independent measurements on simple paramagnetic salts of these elements. The explanation of the discrepancy must, however, be quite different in the two cases, since the ferrites are, at least to a first approximation, non-conducting dielectrics with basically ionic crystal binding. The experimental evidence is unhelpful, particularly, as will be seen below, in the case of copper and magnesium ferrites : initial guidance for the ferrites in general was provided by theoretical considerations of the sign of the expected exchange interaction between the 3d shell electrons of nearest and next-nearest transition-group ions in the ferrite lattice (Verwey, Haayman and Romeijn 1947). The most illuminating evidence has been provided by magnetic saturation measurements on several ranges of mixed ferrites, in which an increasing proportion of the paramagnetic ions are replaced by diamagnetic zinc ions (Gorter 1950a, b). The resulting mixed ferrites are homogeneous and retain the spinel structure, and x-ray analysis shows that the zinc ions go into tetrahedral sites, just as in pure zinc ferrite. Very recently the neutron diffraction technique has been used to reveal information about the magnetic structure of magnetite (Shull 1951, Shull, Wollan and Strauser 1951) and has confirmed Néel's theory, described in the following paragraphs. This technique promises to become one of the most powerful yet applied to the problem of ferromagnetic (and antiferromagnetic) structure.

Meanwhile a comprehensive theory of the magnetic properties of ferrites has been developed by Néel (1948), and Néel's suggestion that a 'superexchange' spin coupling involving intermediate oxygen ions must be invoked for ferrites has been studied by Anderson (1950) for the somewhat simpler case of the antiferromagnetic oxide MnO.

In Néel's theory no particular mechanism is postulated for the coupling between the spins of the various paramagnetic ions : this coupling is represented by equivalent Weiss molecular magnetic fields. The magnetic environment of any chosen magnetic ion consists of a superposition of two groups of magnetic ions, the interaction between each group and the chosen ion being of arbitrary magnitude and sign. In the first approximation, only the nearest neighbours of each group are considered. The groups comprise the tetrahedral ('A-site') and octahedral ('B-site') neighbours respectively. In the first place Néel considers a simple ferrite of the general formula XY_2O_4 in which the X ions are diamagnetic and the Y ions all consist of Fe^{3+}. The case where the X ions also have a permanent magnetic moment is more involved, and that in which the X ions or the Y ions (or both) each comprise more than one element is even more complex. These cases are of practical interest and will be considered later. An outline of the simple theory is given in the Appendix. Figure 3 shows the variety of characteristics possible according to the simple theory.

When it is recalled that this theory applies to the case of a spinel having only one type of magnetic ion (ions of the same element in different valencies or electronic states counting as different types), it will be clear that a general theory for a typical ferrite containing two types of magnetic ion distributed over the A and B sites

would, even on the same phenomenological basis as before, become somewhat intractable. The number of interactions between magnetic ions increases from 3 to 10, and the number of variables of experimental relevance increases further, eqns. (1) and (2) of the Appendix being replaced by four equations each with four terms, together with auxiliary relations between the coefficients. It is, therefore, perhaps a little surprising to find that the observed magnetic saturation and susceptibility characteristics of the majority of ferrites can be tolerably well understood in terms of the simpler version of the theory.

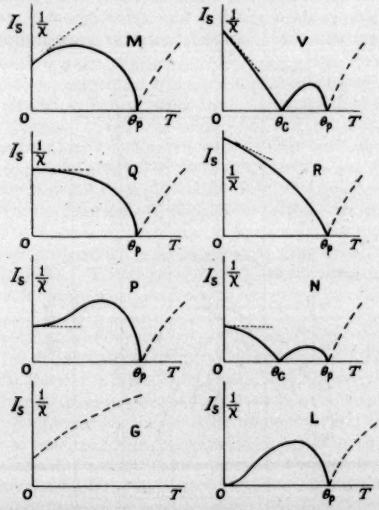


Fig. 3. Theoretical temperature variations of magnetization and paramagnetic susceptibility of a simple ferrimagnetic (*after* Néel 1948, with acknowledgments to the author and to the publishers of *Annales de Physique*).

Paramagnetic susceptibility measurements on a number of simple ferrites and on some mixed ferrites have given general agreement with eqn. (6) (Appendix) and values of $1/\chi_0$ and C have been obtained (Néel 1948). The extrapolated paramagnetic Curie temperature θ_p has been found to be 40–100° above the ferromagnetic Curie point θ_f where this can be determined. The experimental value of C required in eqn. (6) has also been found to be 10–100% greater than that expected from the original theory (Fallot and Maroni 1951). By postulating a temperature coefficient for the molecular field coefficient n in eqns. (1) and (2) so that instead of constant n we have

$$n = n_0(1+\gamma T), \qquad \ldots\ldots(14)$$

Néel (1951 a) has shown that the discrepancy in C can be accounted for with values of γ of the order of -10^{-4}, which are not inconsistent with linear thermal expansion coefficients of the order of 10^{-5} per degree Kelvin, implying a Weiss molecular

field varying as about the inverse tenth power of the ionic separation. According to Pauthenet and Bochirol (1951), the difference between extrapolated θ_p and observed θ_f is also greatly reduced by taking account of the relation (14), and, in fact, the complete magnetization–temperature characteristic below θ_f was shown by the latter authors to be then in tolerably good agreement with theoretical prediction for Co ferrite and magnetite, and somewhat less good for Ni ferrite.

Extensive magnetic saturation measurements have been made on mixed ferrites by Gorter and Guillaud. Gorter (1950 a, b) has investigated the variation of the spontaneous magnetization of mixed ferrites of Mn, Fe, Co, Ni, Cu, Mg and Li with Zn over a range of compositions from zero to about 75 atomic per cent Zn, with the results shown in fig. 4. In this figure, the magnetization is measured in

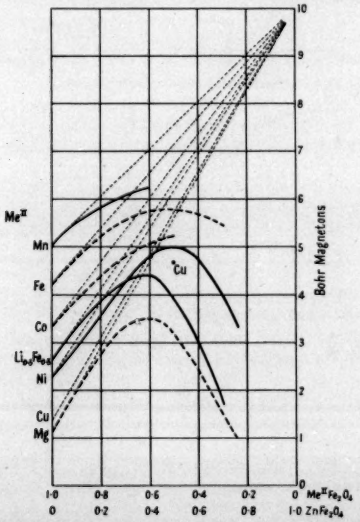


Fig. 4. Variations of saturation magnetization with zinc ferrite content for compound ferrites (*after* Gorter 1950 b, with acknowledgments to the author and to the publishers of *Nature*).

Bohr magnetons per molecule of the mixed ferrite $M_xZn_{1-x}Fe_2O_4$, where M is the bivalent metal other than Zn and x is the fractional atomic proportion of this metal. As the Bohr magneton value represents the net number of parallel electron spins, it is seen that in all cases the magnetization when $x=1$ is considerably less than the value to be expected if all the available spins were parallel. For example, if all the spins in the molecule $NiFe_2O_4$ were parallel, the total spin would be $2+5+5=12$ Bohr magnetons, compared with the observed value of about 2·3. In conjunction with the inference from x-ray structure analysis, that $NiFe_2O_4$ has probably an inversed spinel structure, with half of the Fe^{3+} ions in A sites and half in B sites, and all the Ni^{2+} ions in B sites, this low saturation value is explained by assuming that the A–B interaction favours anti-parallel spin coupling between the A-site

and B-site groups, and is strong enough to overcome any tendency for A–A or B–B couplings to disrupt the parallel alignment of all spins within the A-site or B-site groups separately. On this basis, the spins of the Fe^{3+} ions in A sites would exactly cancel those in B sites, leaving a net value of two Bohr magnetons per molecule arising from the bivalent ions alone. The same argument may be applied to all the simple ferrites MFe_2O_4 when their structure is known to be purely inverse, and predicts integral Bohr magneton values agreeing approximately with, but in general somewhat lower than, the experimental values for Fe, Co and Ni ferrites.

The remaining discrepancy has two possible explanations: the first is as follows. In a free ion, the orbital motion of the unpaired 3d electrons is able to contribute a magnetic moment additional to that arising from their spin. An ion forming part of a solid is subject to a strong electric field of non-spherical symmetry by its neighbour ions in the crystal lattice, and this electric field is, in most cases, able to freeze or 'quench' the orbital motion and thus prevent its contributing any observable magnetic moment. There is evidence from mechanical gyromagnetic and from radio-frequency precession resonance experiments, that the residual orbital contribution may not be negligible for Fe^{2+}, Co^{2+} and Ni^{2+}, and it has been suggested (Gorter 1950a, b, Guillaud 1951a) that this contribution may explain the whole of the observed 'excess' magnetization for the simple ferrites. The Mn^{2+} ion has zero orbital magnetic contribution, even in the absence of crystalline electric fields. The close agreement between the measured and the theoretical 'spin only' magnetizations for Mn ferrite therefore supports the above theory.

The second explanation of the non-integral Bohr magneton values is contained in Néel's theory of ferrimagnetism : the ferrite structure may not be purely inverse. Thus the transfer in $NiFe_2O_4$ of a fraction of the Ni^{2+} from B sites to A sites, with a compensating transfer of Fe^{3+} ions from A sites to B sites, will increase the total B-site spin by $5-2=3$ Bohr magnetons, and reduce the A-site spin by the same amount, per Ni^{2+} ion transferred. It follows that only 5% of the Ni^{2+} ions need be in 'normal' lattice positions to account for the observed excess magnetization of 0·3 Bohr magneton, and a 5% reversion to normal spinel structure could not be detected by available x-ray techniques.

Referring again to fig. 4 we notice that the Bohr magneton value in all cases increases as Zn^{2+} ions are substituted for more and more of the original bivalent ions. The Zn^{2+} ion has a full 3d shell and is thus diamagnetic. The preceding argument therefore shows that simple substitution of, for example, Ni^{2+} ions by Zn^{2+} ions, must lead to a decrease in net Bohr magneton value towards zero for complete substitution, in contradiction to the magnetic evidence. The x-ray evidence, however (Verwey and Heilmann 1947), shows that pure Zn ferrite as ordinarily prepared has a normal spinel structure. The mixed Ni–Zn ferrite structure may, therefore, be assumed to show a gradual transition, as Zn is substituted for Ni, from nearly inverse to essentially normal structure. Hence each Zn^{2+} ion added goes into an A site, displacing an Fe^{3+} ion from there into a B site, this Fe^{3+} ion in turn displacing a Ni^{2+} ion from the crystal altogether (assuming that the mixed ferrite has the desired stoichiometric composition). Thus, if the relative spin alignments in the A and B sites remain unchanged as the Zn substitution proceeds, we shall expect the net spontaneous magnetization to tend towards ten Bohr magnetons per molecule as given by a hypothetical pure

$ZnFe_2O_4$ normal spinel with all the Fe^{3+} spins parallel. It is seen from fig. 4 that in fact the initial tendency for small fractional substitutions by Zn is towards ten Bohr magnetons in all cases. But, at high substitution fractions, the magnetization falls again and, furthermore, it is known that pure Zn ferrite of normal spinel structure is paramagnetic down to the lowest temperatures (Guillaud and Sage 1950). This behaviour shows that the B–B interactions are negative, tending towards an antiferromagnetic arrangement of the spins within the B-site group, which is only broken down when the A sites are sufficiently occupied to enable the stronger (but still negative) A–B interaction to take control, as already seen, for example, in the pure Ni ferrite.

The general form of the curves of fig. 4 is thus accounted for. The so-called Li ferrite has the 'average' molecular formula $(Li_{0\cdot5}Fe_{0\cdot5})Fe_2O_4$ and an inverse spinel crystal structure with a theoretical magnetization of 2·5 magnetons per molecule, agreeing well with the experimental value. The cases of Cu and Mg ferrites, however, show marked discrepancies depending systematically on the heat treatment of the samples. These have been investigated experimentally by Guillaud (1951 c) and by Pauthenet and Bochirol (1951) on the basis of a theory put forward by Néel (1950 a). The theory is simply an extension of the concept of partial reversion to normal spinel structure: reversion is assumed to require an activation energy defined in terms of a temperature θ. The fraction x of bivalent ions which are in equilibrium in A sites at a temperature T is then given statistically by

$$\frac{x(1+x)}{(1-x)^2} = e^{-\theta/T}. \qquad \ldots\ldots(15)$$

Experimentally, values of x over a range of high temperatures can be inferred from magnetic saturation measurements at low temperature on samples quenched from the high temperatures. By treating θ as an adjustable constant, Pauthenet and Bochirol (1951) have found approximate agreement with the formula (15) for Cu ferrite over the T range of 600–1300° K, but only very rough agreement for Mg ferrite. The magnetization of Bohr magnetons per molecule should of course be of the form

$$m + 2x(5-m) \qquad \ldots\ldots(16)$$

where m is the magneton value for the bivalent ion. If m itself were regarded as an adjustable constant in fitting the measurements of Pauthenet and Bochirol to the law (15), a value of m closer to unity than the theoretical value zero would give better agreement for the Mg ferrite. On the other hand, the experimental results for low values of T (a point below 500° K is shown) may be in error owing to the possibly long time needed for the establishment of equilibrium prior to quenching.

The question of 'correct' interpretation of the spontaneous magnetization of simple ferrites cannot yet be regarded as entirely settled. Several series of measurements on mixed ferrites of varying proportions of Ni–Mn, Co–Ni and Mn–Co–Ni (Guillaud 1951 a) have shown that the Bohr magneton value and the Curie temperature vary linearly with composition between the values found for the simple ferrites. This linearity has been taken as a confirmation that the Bohr magneton values of the simple ferrites include a constant orbital contribution, though it would seem equally consistent with a characteristic partial reversion to normal spinel structure for each component ferrite. The precession resonance measurements, to be discussed later, appear to be the only evidence of any weight in favour of an orbital contribution and that somewhat smaller than was originally thought to be the case (Yager, Merritt and Guillaud 1951).

There seems to be, as yet, no direct evidence (analogous to that provided for the negative B–B interaction by the drooping of the curves in fig. 4 at high Zn fractions) that the A–A interaction is negative. Only Néel's (1948) theory of the shape of the magnetization–temperature characteristic points to this for the relatively few cases in which it has been worked out in the literature (incidentally, Néel's calculations suggest that in magnetite the B–B interaction may be very small or even slightly positive). It may prove to be impossible to obtain the direct evidence desired, on account of the way in which suitable substitution ions appear to influence the occupation of the A sites as well as the B sites. For example, the natural mineral 'spinel' itself, $MgAl_2O_4$, has the 'normal' structure. A mixed 'ferrite' in which some of the Fe^{3+} had been replaced by Al^{3+} might therefore be expected to have had the Fe^{3+} ions in B sites replaced. The x-ray evidence, however, in the case of Mg ferrite (Nicks and Welch 1951) when combined with the magnetic saturation evidence, suggests that the structure remains nearly 'inverse' so far as the Fe and Mg ions are concerned up to at least 50% substitution, that is to the composition $MgAlFeO_4$, while the Al^{3+} seems all to go into the B sites. Magnetic saturation measurements on $MgAlFeO_4$ at 20° K (Jones and Roberts 1951) give a value of 0·3 Bohr magneton per molecule, which can most easily be explained by assuming that more than half of the Mg ions are in A sites and more than half of the Fe^{3+} ions are in B sites. The Curie temperature of $MgAlFeO_4$ is about 230° K.

Substitution of Al^{3+} for Fe^{3+} ions has also been carried out in magnetite : Guillaud and Michel (1951) have found a linear fall of magnetization with up to 10 atomic per cent substitution of the Fe^{3+} ions, the fall tending towards a value of one Bohr magneton per molecule when half of the total number of Fe^{3+} ions have been replaced. This may be interpreted (Guillaud 1951a) in terms of Fe^{3+} substitution of the B site Fe^{3+}, leaving the B site Fe^{2+} anti-parallel to the undisturbed A site Fe^{2+} ion. A second possibility (Guillaud 1951a) is that the Al^{3+} goes into a B site occupied by Fe^{2+} displacing this to an A site and turning an Fe^{3+} ion from here out of the molecule, the normal negative A–B interaction then giving the molecular moment of one Bohr magneton.

§ 5. NATURE OF THE MAGNETIC INTERACTIONS

Verwey (1947) and, in greater detail, Néel (1948) have examined the possibility that the spin–spin interactions responsible for the ferromagnetism (or ferrimagnetism) of the ferrites might be of the same quantum-mechanical 'exchange' type as that considered in the case of the ferromagnetic metals. As is well known, this exchange interaction is usually supposed to depend, in magnitude and sign, on the extent and form of the overlap of the 3d-electron wave functions of neighbouring ions. The interaction favours anti-parallel spins when the ions are nearer together than a critical distance, and parallel spins for greater separations, but the strength of the interaction falls off rapidly for large separations. In a typical spinel structure with lattice unit cell cube edge of 8·40 Å, the distances between the relevant ion centre positions are : A–A separation = 3·65 Å; A–B separation = 3·48 Å; B–B separation = 2·97 Å. These distances are for the nearest neighbours in each case. In metallic α-iron (the normal ferromagnetic form) the nearest neighbour separation is 2·86 Å between centres, that is, less than all the magnetic ion separations in the ferrite structure. Yet, while the interactions are undoubtedly positive (in Néel's sense) in the metal, and would therefore remain positive though weaker at the larger

separations in the ferrite, there is no doubt that, in fact, the A–B interaction, and probably also the B–B and A–A interactions, are negative in the ferrites.

To be rigorous in this argument, one should take account of the mean diameters of the 3d-electron shells of the magnetic ions concerned. When, however, one also notes the presence of the much larger outer electron shells of the intervening O^{2-} ions in the ferrites, it becomes doubtful whether there can be any significant overlap of the 3d shells of any of the magnetic ions, and the fact that the ferrites are other than paramagnetic appears inexplicable. The same conclusion must be reached in the case of a number of other compounds such as MnO, MnS which are known to be purely antiferromagnetic (Néel 1948).

A way out of the difficulty is provided by Kramers' theory of 'superexchange' interaction (1934). This theory has been discussed and simplified somewhat by Anderson (1950) and Van Vleck (1951a) for the case of pure antiferromagnetism. A second mechanism, which may account for the weak positive B–B interaction in magnetite and for most of the ferromagnetism of certain compounds having the perovskite structure, is that of 'double exchange' described recently by Zener (1951a, b): this will be discussed later.

Superexchange provides spin coupling between the magnetic ions by way of excited electronic states of the intermediate oxygen ions, so that the latter act as 'relays' instead of passive obstacles. Consider the working of the superexchange mechanism in, for example, the A–B interaction in Ni ferrite, and assume initially that the crystalline binding forces are ionic in nature.

The normal state of each oxygen ion will be O^{2-} with an outer electron configuration $2p^6$; each nickel ion will be Ni^{2+} with $3d^8$ (all assumed to be in B sites), while the iron ions will be Fe^{3+} with $3d^5$ (filling all the A sites and half the B sites). The O^{2-} ion in its normal state has thus a full 2p shell and is, therefore, diamagnetic. If now an O^{2-} ion is excited, one of its 2p electrons might move out to a 3s shell, but this is unlikely because of the greater diameter involved relative to the already quite close packing of the oxygen ions. A more likely occurrence is for the 2p electron to leave the oxygen ion entirely and to attach itself to probably the 3d shell of one of the neighbouring (and much smaller) metal ions. Without going into the relative weighting between the A-site Fe^{3+}, the B-site Fe^{3+} and the B-site Ni^{2+} ions, we can say that the spin of the 2p electron involved will be opposite to the resultant spin of the metal ion, whichever of the latter is involved (since all have at least half-full 3d shells). Consequently the oxygen ion becomes O^- with an outer configuration $2p^5$ and a resultant spin moment of one Bohr magneton *parallel* to the resultant spin of the affected metal ion. The O^- ion, in contrast with the O^{2-} ion, is thus itself paramagnetic, and is, therefore, able to interact by the ordinary exchange mechanism with neighbouring magnetic ions other than the particular metal ion to which it has lost an electron. It may be noted at this point that the resultant total spin of the latter metal ion and the O^- ion taken together has the same value as that possessed by the metal ion alone before the electron transfer. The externally observable Bohr magneton value per molecule is, therefore, unaltered by the transfer mechanism described.

The centre-to-centre separations between oxygens and nearest metal ions are about 1·9 Å for A sites and about 2·1 Å for B sites, so that the overlap of the unpaired 2p electron of the O^- with the 3d shell of an adjacent metal ion in its normal state is probably sufficient to align the spins of these two ions anti-parallel. Taking this ordinary exchange coupling together with the electron transfer coupling, we infer

that the overall 'superexchange' coupling between metal ions separated by an oxygen ion tends to align the spins of the metal ions anti-parallel, in agreement with Néel's ferrimagnetic model.

The preceding outline of the superexchange mechanism has been reduced to the barest essentials. It must be observed that in a uniform crystal, the permanent transfer of an electron from one particular O^{2-} ion to any particular metal ion neighbour is almost certainly not an equilibrium condition. The actual condition is more akin to the dynamical equilibrium of covalent bonds superimposed on the basic ionic binding. Furthermore, as Anderson (1950) has discussed, the electron which may be considered (in an initial analysis) as transferring has some probability of going to a 4s shell rather than to a 3d shell on the metal ion, because the free atoms of Mn, Fe, Co and Ni all have full $4s^2$ configurations with 'holes' left in the 3d shell, and the ions may be expected to show some sign of this preference even in the crystal environment. This introduces a further link in the superexchange chain: unless there is spin coupling (originating presumably from ordinary exchange in the perturbed environment) between the 3d and 4s shells in the affected metal ion, there can be no superexchange in this case. The sign of the 3d-to-4s exchange will enter into that of the superexchange.

Besides the possibility just noted, there is one that was earlier dismissed as improbable. This is that the excited electron from the O^{2-} ion remains on that ion in a 3s shell. A somewhat analogous possibility (Van Vleck 1951) is that, in consequence of a 2p electron leaving the oxygen ion entirely, a 2s electron in the same ion has some chance of moving out to fill the resultant 'hole' in the 2p shell, leaving a hole in the 2s shell. The significance of these two possibilities is that the magnetically important unpaired electron left on the oxygen ion in either case is in an s shell, and its orbital wave function therefore has spherical symmetry. This is to be contrasted with the dumb-bell shaped wave function of the unpaired electron in the O^- 2p shell, the state believed to be most probable in the case of simple antiferromagnetic structures like MnO. Strong evidence for the latter belief is provided by the neutron diffraction results (Shull and Smart 1949, Anderson 1950) which show that in MnO, the Mn^{2+} ions which are next nearest neighbours are arrayed in a magnetic super-lattice with alternately anti-parallel spins, but that nearest Mn^{2+} neighbours are uncorrelated as regards their spin directions.

Brockman (1951) has briefly considered the more involved symmetry of the ferrite structure. The angle subtended by two adjacent B sites at the centres of each of the two neighbouring oxygen ions is 90°. In the case of two adjacent A sites, each of the three intervening oxygen ions subtends an angle of about 80°. There is only one shortest superexchange path from an A site to a B site, and the oxygen ion in this case subtends an angle of about 125°. Thus, if the 2p configuration were the only significant one for the excited oxygen ions, we should expect very weak B–B interaction, weak A–A interaction, and relatively strong A–B interaction. The slight deviations of the metal ions from positions of ideal symmetry, revealed by x-ray diffraction, should not alter this order of magnitudes. As already seen, the magnetic evidence (fig. 4) indicates that the B–B interaction is important in the magnetically diluted ferrites and suggests, therefore, that there may be a considerable component of 2s and 3s electron configuration in the excited oxygen ions.

Zener's (1951 a, b) double exchange mechanism involves the simultaneous transfer of an electron from one metal ion to a neighbouring oxygen ion, and of

another electron from the same oxygen ion to a second metal ion neighbour, with a preference, as in superexchange, for diametrically opposite metal ions, if the 2p shell of the oxygen is that chiefly involved. Double-exchange is appreciably probable at ordinary temperatures (on account of the activation energy required) only when the metal ions concerned are simply ions of the same element in different valencies. Zener has discussed the case of Mn^{3+} and Mn^{4+} ions in mixed perovskite-type crystals, which have been investigated experimentally by Jonker and van Santen (1950). It seems likely that the same double-exchange mechanism may be at work in addition to a superexchange in the case of magnetite, which contains Fe^{2+} and Fe^{3+} ions in similar lattice positions (the B sites).

As Zener shows, double-exchange will be associated quite generally with parallel spin alignment forces between the affected metal ions. Thus, in the case of magnetite, double-exchange will be acting in opposition to superexchange in the B–B interaction, and may account for Néel's (1948) result that this interaction is slightly positive for this material.

§ 6. PERMEABILITY–FREQUENCY BEHAVIOUR

The anomalous magnetic behaviour of ferrites at radio frequencies appears to have been noted first by Snoek (1946), who also proposed (1947) to account for the phenomenon by means of the gyromagnetic precession resonance mechanism discussed theoretically many years earlier by Landau and Lifshitz (1935). A number of measurements, at frequencies up to 10^{10} c/s on ferrites and on γ-Fe_2O_3, all in powder form dispersed in a non-magnetic medium, have been made by Birks (1947, 1948 a, c) and interpreted in the same light. Birks employed a semi-empirical formula to extrapolate from his measurements on samples at various dilutions to the state of zero dilution, and implied that such extrapolated results represented the magnetic behaviour of the solid ferrite material. It is questionable whether extrapolation in this manner can give results corresponding to the behaviour of a large single crystal sample, but it probably does represent approximately that of the more usual polycrystalline sintered material, in which the originally separate grains retain to some extent their individual identities.

Measurements on pressed powder samples of γ-Fe_2O_3 and on sintered samples of Ni ferrites, Cu–Pb ferrites, Mg ferrites and Cu–Zn ferrites have been reported by Flegler (1950). These measurements extended up to 3 000 Mc/s, though at rather widely spaced frequencies, and all showed apparently a steady fall of permeability with increase of frequency. More recent measurements reported by Welch, Nicks, Fairweather and Roberts (1950), were made at rather closer frequencies in the region of 1 000 Mc/s and showed that there was a subsidiary peak of permeability in this neighbourhood for a range of Mg ferrites in pressed-powder form. Essentially similar behaviour was found for a sintered Mg ferrite by Rado, Wright and Emerson (1950). The results found by Roberts (1951b) at room temperature are shown in fig. 5. Measurements at liquid oxygen temperature on these same Mg ferrites show: (*a*) a marked reduction in permeability up to about 200 Mc/s relative to the room temperature values, and (*b*) a considerable rise in the frequency, and some rise in the magnitude, of the subsidiary peak in permeability.

The magnetic behaviour of magnetite at low temperature shows an anomaly even at low frequency (Snoek 1947–9): the permeability rises to a sharp

peak at about 130° K, falling to a very low value at lower temperatures. The behaviour is thus analogous to an 'inverse Curie point' effect, and appears to be associated with the ordering of Fe^{2+} and Fe^{3+} ions into alternate B sites along (110) directions (Verwey and de Boer 1936). Bickford (1949 b, 1950) and Domenicali (1950) have concluded that the ordering takes place in two stages, the magnetic anisotropy being like that of metallic Ni down to 130° K, like that of Fe between 130° K and 110° K, and somewhat like that of Co below 110° K. The electrical conductivity and specific heat also pass through transitions at about 110° K.

Single crystals of Ni ferrite have been grown by Galt, Matthias and Remeika (1950), and their permeability measured over a range of temperatures and frequencies. At room temperature the permeability falls from 80 at 50 kc/s to about 45 at 2 Mc/s, while at 83° K it has fallen to about 15 throughout this frequency range. Snoek (1947–9) had previously found a generally similar behaviour

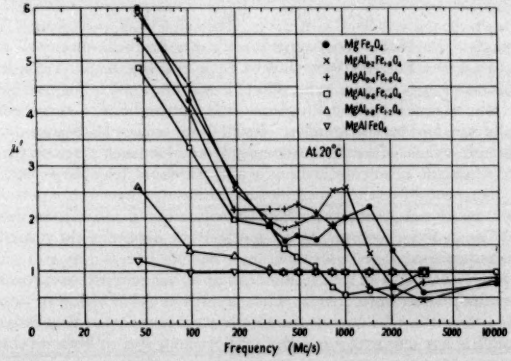


Fig. 5. Dispersion of ferromagnetic permeability of compound magnesium ferrite-aluminates. (Roberts 1951 b, with acknowledgments to the publishers of *Journal de Physique et le Radium*.)

for a polycrystalline sample of Cu–Zn ferrite over the frequency range of approximately 1–20 Mc/s and the temperature range from 93° K to 293° K.

The interpretation and correlation of the above and similar evidence within a single comprehensive theory does not as yet appear to be possible, but several probable features of such a theory are beginning to be clear.

The following mechanisms have been considered in the literature: (i) classical eddy currents in possibly inhomogeneous material; (ii) simple electronic precession resonance in crystalline magnetic anisotropy field, and granular shape anisotropy field in inhomogeneous material; (iii) more complex precession resonance in combination of anisotropy and ferrimagnetic exchange fields; (iv) domain wall resonance; (v) electromagnetic 'body resonance'. Flegler (1950) has examined in some detail the possibility that, in spite of the relatively low d.c. conductivity, eddy currents within the crystallites of a sintered ferrite might account for the observed general fall of permeability at high frequency. If the sintered material is regarded as magnetically inhomogeneous, with high permeability as well as

high conductivity within each crystallite, a simple eddy current mechanism, modified by Becker's (1938, 1939) and/or Kittel's (1946, 1948) additional domain-wall eddy-current damping effects, may account for much of the fall of permeability in some cases. In general, however, the observed permeability of most ferrites falls at too low a frequency, and at too fast a rate with increase of frequency, to be explained wholly by an eddy current mechanism.

Snoek has shown (1947) that, for a range of ferrites of different low-frequency permeability, but of the same saturation magnetization, the frequency at which the permeability has fallen to half its low-frequency value varies nearly inversely as the low-frequency permeability.

Corresponding to Snoek's relationship, Polder has shown (1950) that the frequency at which the tangent of the magnetic loss angle, $\tan\delta_\mu$, has a particular value (0·1) for a range of similar ferrites varies inversely as the low-frequency permeability, when the complex permeability is written as

$$\mu=\mu'-j\mu'', \qquad \tan\delta_\mu=\mu''/\mu'. \qquad \ldots\ldots(17)$$

The mechanism of the gyromagnetic resonance has been discussed in detail in another paper by Polder (1949), and by Snoek (1948), Kittel (1951) and Van Vleck (1951 a).

This resonance effect in its simple form is not peculiar to ferrites but can occur in all ferro- and para-magnetics. It will not therefore be discussed in detail here, although it has generally been considered that such a resonance in the crystalline magnetic anisotropy field probably accounts for a large part of the magnetic dispersion of ferrites, thereby suggesting that most of the permeability arises from rotational rather than from domain wall displacement processes. It does not seem able, however, to account for the details of the dispersion at very high frequencies.

A more complex type of precession resonance has recently been considered for antiferromagnetics and ferrimagnetics by Kittel (1951 b). The treatment has formal similarities with that of the simple resonance mechanism for the case where sample shape anisotropy is taken into account, for, in both mechanisms, the instantaneous torque tending to sustain the precession may be regarded as varying through the precession cycle. In the case of a monocrystalline spherical sample of a ferrimagnetic ferrite immersed in an external field H_0 along a preferred direction of magnetization, Kittel obtains for the possible precession resonance frequencies, on the basis of pure spin precession:

$$\nu_0=2{\cdot}80\{H_0-aH_E\pm[H_A(H_A+2H_E)+aH_E(aH_E-2H_A)]^{1/2}\}\ \text{Mc/s}, \qquad \ldots\ldots(18)$$

where H_A = crystalline magnetic anisotropy field, in oersteds, assumed to be the same for spins in all relevant lattice sites, H_E = Weiss molecular field is oersteds acting on spins in (say) A sites due to a full complement of spins in B sites, or vice versa, and $2a$ = ratio of the net static magnetization to the greater of the two sub-lattice (A-site or B-site) magnetizations. In the limiting case of a pure ferromagnetic, where $2a=1$, the above result reduces, taking the positive root, to

$$\nu_0=2{\cdot}80(H_0+H_A)\ \text{Mc/s}, \qquad \ldots\ldots(19)$$

just as in the elementary theory. The interest of the new theory is that, for plausible values of H_A and H_E in ferrites, it leads to notably larger values of ν_0 than the simple theory, and may explain some of the anomalous gyromagnetic

g-factors, besides the dispersion behaviour, found for ferrites. The new theory is evidently over-simplified in comparison with Néel's description of the magnetic structure of ferrites, and it is to be expected in general that there may be two or more resonance frequencies within, or near to, the presently accessible range of radio-frequency measurements.

Two other possible causes of a resonance behaviour in the observed permeability within the presently accessible frequency range must next be considered. One is specifically a domain wall effect, first considered by Döring (1948) and discussed by Becker (1951). The transient spin precessions involved in a moving wall set up a magnetic double layer in the wall, which thereby acquires energy additional to its static value. Döring represents this as kinetic energy, and by associating the corresponding 'mass' with the elastic restoring force determining the wall displacement susceptibility under quasi-static conditions, obtains a wall resonance frequency:

$$\omega_0 = I_s(\pi t/\chi_0 l)^{1/2}(e/mc), \qquad \ldots\ldots(20)$$

where I_s = saturation magnetization, χ_0 = initial low frequency displacement susceptibility, t = wall thickness, l = average domain thickness. Insertion of typical data in (20) shows that ω_0 is considerably lower than the domain wall relaxation frequency due to Becker's (1939) eddy current damping mechanism, even in the case of the ferromagnetic metals. The above domain wall resonance frequency apparently varies more slowly with susceptibility than the purely rotational resonance mechanism considered earlier, but the explicit relationship in (20) is more complex since the forces controlling χ_0 also in general influence t and l.

There appears to be as yet little experimental evidence in any magnetic material for a true domain wall resonance as distinct from a pure rotational resonance, though it is possible that some of the observed characteristics hitherto ascribed to pure rotation may be interpreted instead as due to wall effects.

Another possible cause of resonance in the apparent permeability variation with frequency is to be found in the so-called 'body-resonance' effect: this effect has been discussed by Brockman (1949). It arises from the possibility that the wavelength of plane electromagnetic waves propagating within a ferrite or similar material may, on account of the combination of high permeability and high permittivity in the material, become comparable at the frequency of measurement with the physical dimensions of the specimen employed for the permeability measurements. The general theory of guided waves and cavity resonators makes it clear that, under certain conditions as to the losses within the bulk of the sample and the wave reflection coefficient at its boundaries, the specimen as a whole may behave like a cavity resonator of relatively high Q-factor. The theory also shows how the specimen may be subdivided so as to increase the frequencies of any such 'body' resonances above the desired working range, without appreciably reducing the low-frequency permeability effective for applied magnetic fields of the desired form (e.g. toroidal). Brockman's evidence (1949) shows clearly that incipient body resonance can be important in some cases. But in general we may regard body resonance as much less significant than would appear at first sight, for the following reasons: (*a*) the dielectric permittivity, although perhaps of the order of 10^5 at frequencies of 100 c/s and lower, invariably falls more or less rapidly with increase of frequency, as will be seen in § 9, (*b*) the permeability, as already seen, falls with increasing frequency and commences to

do so at a frequency which decreases with increase of the low-frequency permeability, and (*c*) both the dielectric and magnetic loss angles are large over the range of otherwise possible body resonance effects. Consequently, the simple estimation of possible body resonance frequencies from the known low-frequency permeability and permittivity of the material may give values too small by a factor of 2 or more, and the effect of any such resonance will be strongly damped by the inherent losses in the material in the frequency range of the resonance. Nevertheless, one type of body resonance will always ultimately occur in permeability measurements on samples of fixed dimensions over very wide frequency ranges. This arises when the sample thickness in the direction of propagation of the fundamental wave mode used for the measurements becomes comparable with one quarter wavelength in the sample material. Any error due to this effect may always be eliminated by making two impedance measurements on the same sample under two known boundary conditions, usually 'open-circuit' and 'short-circuit' (Birks 1948a). It is assumed throughout the present Report that this, or an equivalent, method has been employed to eliminate dielectric effects from all the permeability results discussed.

§ 7. 'DAMPING'

We have seen how both of the fundamental magnetization processes, of spin rotation within a single domain, and of domain wall displacement, should have a resonance behaviour. Yet there is very little evidence of a sharp resonance behaviour either in ferrites or in metals sufficient to show a peak of the real part of the permeability at a non-zero frequency. In most ferrites (after allowing where necessary for body resonance effects) the dispersion of permeability is only a little steeper than could be accounted for by a simple relaxation process. Strangely, the sharpest resonance so far found for any material has been reported for a nickel–iron alloy in rod from (Wieberdink and Kronig 1948). We should, of course, expect some damping of the resonance in metals by eddy current effects, though Döring (1948) and Becker (1951) have shown that the domain wall resonance, at least, should not be heavily damped through this cause. In ferrites eddy current damping should be negligible in the frequency range of 1–100 Mc/s where the main resonance seems mostly to occur.

It is possible to have apparent 'damping' without any true energy dissipation. In a material which is not magnetically homogeneous, it is plausible that the resonance frequency ν_0 is, in general, different in different elementary volumes (even when no external static field is applied). A polycrystalline sample in an external static field would, of course, experience different effective orientations of the crystalline magnetic anisotropy fields relative to the applied field. Crystallite shape may have an analogous effect. In each case the result will be a broadening of the single resonance into a band determined by the spread of the effective internal fields, so giving an apparent 'damping' of an assumed single resonance.

The above effect can be eliminated by the use of single crystal samples. It is found (Galt, Matthias and Remeika 1950) that even single crystals of ferrite give a very broad dispersion of permeability, with little indication of resonance for that part of the permeability which is due to domain wall displacement. The rotational part does not seem to have been explored for a possible sharp natural resonance in single crystals, but the induced resonance can be very sharp (Kittel 1951a, Yager, Galt, Merritt and Wood 1950), in contrast to the behaviour of

metals (Kittel 1951 a, Bloembergen 1950). Kittel (1951 a) and Van Vleck (1951 b) have considered possible causes for the damping, and work in this direction has been published by Akhieser (1946), Katayama (1948) and Huzimura (1950). It cannot be said yet that there is any satisfactory explanation of the domain wall damping in ferrites. There is so far only rather limited evidence concerning the rotational damping. From one point of view it would seem (somewhat paradoxically) desirable that precession should be damped in order that domain wall displacement might proceed with less restriction from Döring's (1948) resonance mechanism. Perhaps the strong rotational damping found in metals by Bloembergen (1950) accounts for the absence of detectable domain wall resonance in most metals, and for the relatively small magnetic dispersion up to 1000 Mc/s in most cases. Perhaps also this damping is due in some way to the presence of the conduction electrons (it is noteworthy that the magnetic dispersion does not set in in magnetite, the most strongly conducting ferrite, until rather higher frequencies than in the more highly insulating ferrites, but this observation requires confirmation by more extensive measurements to eliminate the effect of anisotropy fields).

Finally, in those cases where incipient body resonance is significant, 'apparent' magnetic damping will arise by way of the purely dielectric losses in the equivalent resonator.

§8. MAGNETIC LOSSES AT LOW FREQUENCIES AND LOW LEVELS

The 'damping' considered in the preceding paragraphs is that required to account for the absence in general of sharp resonances in the observed magnetic dispersion over frequencies from 1 Mc/s to 10^4 Mc/s or higher. We shall here consider briefly the losses that become of practical interest only at frequencies lower than the lowest major resonance frequency of the material, and so below 10 Mc/s for most materials of present interest. The hysteresis losses with a.c. magnetizations sufficient to cause saturation may be expected to behave essentially as for metals at the frequencies considered here, and will not be discussed further. The hysteresis, 'eddy-current' and residual losses in the Rayleigh loop region of magnetization do, however, show some interesting and perhaps unexpected characteristics. The methods of measuring and separating the latter loss components have been described by Legg (1936) and more recently by Richards, Buckley, Bardell and Lynch (1950), who have also given some experimental results for various ferrites, together with results for insulated iron and iron alloy powder material. Perusal of the latter results reveals that the hysteresis coefficient h, the 'residual' coefficient c and the 'eddy current' coefficient e are in general quite comparable for ferrites and the insulated metal powders. Neglecting the possibility of some conduction within crystallites being masked in measurements of bulk resistivity by inter-crystalline insulation in ferrites, it does not seem possible to account for the observed 'eddy current' loss in terms of the relatively large sample cross section. It seems more likely that the quantity measured as an 'eddy current' loss in ferrites arises in that part of the low-frequency 'tail' of a high-frequency damped resonance or relaxation behaviour which gives a magnetic loss angle proportional to frequency at constant induction. Legg's formula, which assumes c to be independent of frequency, is an empirical approximation. Richter (1937) has shown that in some alloys c varies with frequency, and the high-frequency loss in ferrites is best explained as a frequency-dependent residual loss if Legg's representation is used (Snoek 1947–9). A comprehensive

physical interpretation of deviations from Legg's formula is provided for both the hysteresis and residual losses by Néel's theory (1950 c) of these losses in magnetic materials in general. It appears that in ferrites the distributions in amplitude and in space (in the crystal lattice) of the elementary magnetic 'barriers', due presumably to dislocations, vacancies or foreign ions in the lattice, is markedly different from those in the ferromagnetic metals and alloys. This is not, of course, surprising in retrospect, in view of the very different physical structures. Some correlation between structural defects and the magnetic losses has been discussed by Snoek (1947–9) who found that shortage of oxygen in the furnace atmosphere during the preparation of a ferrite would adversely affect the properties. The correlation is, however, not simple, and requires detailed investigation for each particular ferrite. Kingsnorth (1950) has given considerable relevant information for the case of Mg ferrite, which may be summarized as follows: (i) in general, all the low-frequency magnetic losses (expressed as h, c and e) in quenched samples, and c in slowly cooled samples, increase steeply, and the low-frequency permeability μ decreases steeply with increasing oxygen defect in all cases; (ii) in general h and c are somewhat lower for quenched than for slowly cooled samples of the same oxygen content, and the quenched samples have somewhat higher low-frequency permeability μ.

It is clear how the overriding sensitivity of μ itself to oxygen defect may mask the variations of magnetic loss angle due also to this cause: there is scope for much further systematic work in this field. Some interesting data on the variation of the losses with composition for mixed Ni–Zn ferrites have been given by Harvey, Hegyi and Leverenz (1950), who found, *inter alia*, that the apparent eddy current loss was almost independent of the d.c. resistivity.

Snoek's work (1947–9) showed how the oxygen uptake, and hence the desirable magnetic properties, could be improved by sintering procedures which maintain a porous texture as long as possible. The poor magnetic properties of the oxygen starved ferrites probably arises, at least in part, from an excess of magnetite caused by the breakdown of the ferrite, as seen in § 3 (v). Magnetite has an exceptionally large magnetostriction, which will set up stresses in the ferrite lattice, increasing the coercive force and reducing the low-frequency permeability. A distribution throughout the ferrite lattice of 'inclusions' of magnetite may thus be the cause of the magnetic 'barriers' required in Néel's (1950) theory.

The slow migration and intensification of such 'barriers' following the sudden application of a small magnetic field appears to be the origin of the long time scale 'after-effect' phenomena discussed by Snoek (1947–9) for certain non-stoichiometric ferrites.

A closely associated mechanism is probably the explanation of a marked dielectric after-effect also found in some ferrites, as will be seen in the following Section.

§ 9. DIELECTRIC AND SEMICONDUCTING PROPERTIES

(i) *Engineering Considerations*

Partly on account of their less immedate practical interest and partly because of their intrinsic complexity, much less information is generally available on the dielectric properties than on the magnetic properties of ferrites. The main interest of the ferrites is as magnetic materials, and the dielectric characteristics

are, in general, relevant only in so far as they may interfere with the full development of the magnetic qualities. An exception to this general statement arises in the special cases in which wave propagation through the bulk of the material is important, as, for example, in certain conceivable radio lens arrangements for directional aerials for very short wavelengths.

In the more usual application of ferrites, as cores for transformers and inductors, it was at first considered sufficient if the electrical conductivity of the material, as measured in the sintered polycrystalline form in which it is used, were small enough to ensure that classical eddy current effects are negligible. In this approximation the displacement current in the ferrite is regarded as unimportant compared with the conduction current, so that the eddy current effect is then negligible if the 'skin depth' d is much larger than the core cross-sectional dimensions transverse to the working magnetic flux, where

$$d = (10^9/f\mu\sigma)^{1/2}/2\pi \text{ cm}, \qquad \ldots\ldots(21)$$

f = frequency in c/s, μ = permeability of ferrite relative to vacuum, σ = conductivity of ferrite in mho/cm. Thus for a typical Mn–Zn ferrite, having $\mu \simeq 1000$ and

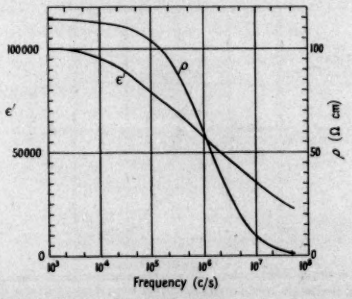


Fig. 6. Dispersion of dielectric permittivity of manganese–zinc ferrite (*after* Polder 1950, with acknowledgments to the author and to the Institution of Electrical Engineers).

d.c. $\sigma \simeq 10^{-2}$ mho/cm, we find for a frequency of 4 Mc/s that $d \simeq 0{\cdot}8$ cm, so that eddy currents are beginning to be significant for a transformer core of the dimensions that would be used at such frequencies, even if no other adverse effect were present. On the other hand, Ni–Zn ferrite, having d.c. $\sigma \simeq 10^{-6}$ mho/cm or less, should have entirely negligible eddy currents under any practical conditions.

There are, however, two adverse effects, rather intimately connected, which may invalidate the above 'skin depth' calculation. These will be apparent from the measured permittivity and resistivity variations with frequency, shown for Mn–Zn and Ni–Zn ferrites in figs. 6 and 7 (taken from Polder 1950). We see that (*a*) since the resistivity falls (σ rises) with increase of frequency, the 'skin depth' d falls more rapidly than $f^{-1/2}$, so that the classical skin effect is always more serious than estimated from the d.c. conductivity; (*b*) the real part ϵ' of the permittivity (relative to vacuum) may be so large at the lower frequencies that wave-propagation effects can be important in cores of practical dimensions.

The wavelength λ_1, of a plane wave within the material is given by

$$\lambda_1 = \lambda_0/(\mu\epsilon)^{1/2}, \qquad \ldots\ldots(22)$$

where λ_0 is the wavelength in vacuum of plane waves of the same frequency. In the case of Mn–Zn ferrite for example, if we neglect the dielectric and magnetic losses, we have $\epsilon \simeq 40000$ and $\mu \simeq 600$ at 4 Mc/s, so that $\lambda_1 \simeq 1{\cdot}5$ cm. The electric and magnetic fields must necessarily be far from uniform across the core cross section under such conditions. There will be certain core dimensions, relative to λ_1, for which standing waves of large amplitude may be set up within the core material, giving rise to resonance phenomena in the externally observed 'magnetic' characteristics of the transformer or inductor. Detailed analyses of such resonance effects (referred to as 'body resonances') are possible by regarding the core as an electromagnetic 'cavity' resonator as noted in § 6. A preliminary discussion of such resonances has been given by Brockman (1949). As may be expected from the comparable magnitudes of d and λ_1 found above for a particular case, the more exact analysis indicates that while waves can propagate in the core, their rate of attenuation is usually so high that resonances are likely

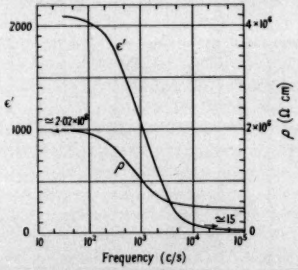


Fig. 7. Dispersion of dielectric permittivity of nickel–zinc ferrite (*after* Polder 1950, with acknowledgments to the author and to the Institution of Electrical Engineers).

to be too heavily damped to be easily detected in the external circuit. We may conclude that the overall result of the usual dielectric characteristics is to increase the apparent magnetic losses of wound cores appreciably more rapidly than classical skin effect would do alone, as estimated from the d.c. conductivity of the material.

(ii) *Dielectric Properties*

It is possible to arrive at an approximate value for the high (optical) frequency permittivity of ferrites by considering the polarizability of their constituent ions. Thus, if we neglect the contributions of ionic displacements (at infra-red and lower frequencies) and their possible reactions on the polarization of the individual ions, the total polarization of the crystal due to a given externally applied field will be given by the sum of the component ionic polarization caused by the corresponding local fields. Now, of the 56 ions comprising the unit cell of the ferrite crystal lattice, 32 are oxygens, and the polarizability α of the O^{2-} ion greatly exceeds that of any of the metal ions occurring in ferrites. Neglecting

the contribution of the metal ions entirely, and assuming the Lorenz–Lorentz local field formula, we obtain for the permittivity

$$\epsilon = \frac{1 + 8\pi N\alpha/3V}{1 - 4\pi N\alpha/3V}, \qquad \ldots\ldots(23)$$

where N = number of ions in volume V. Taking a ferrite unit cell with $V = 8{\cdot}4^3$ cubic Ångströms and $\alpha = 2{\cdot}7$ cubic Ångströms for the O^{2-} ion (Rice 1940), we find $\epsilon \simeq 5{\cdot}7$, which is in good agreement with the measured optical refractive index 2·3–2·4 (Posnjak 1930, Roberts and Merwin 1931). Measurements on sintered polycrystalline ferrites at the highest radio frequencies give higher ϵ values of about 10 (Rado, Wright and Emerson 1950), thus indicating the presence of appreciable dispersion in the infra-red frequency range. Polder's (1950) data for Mn–Zn and Ni–Zn ferrites show (figs. 6 and 7) that at the lower radio and audio frequencies ϵ values rise by factors of the order of $10^2 - 10^4$.

Good crystalline samples of the majority of ferrites appear to have a high degree of cubic symmetry, the crystal structure belonging to the space group $O_h^7 - Fd3m$. Materials with this symmetry cannot be piezoelectric or ferroelectric (ferroelectricity might otherwise be a possible explanation of the very high

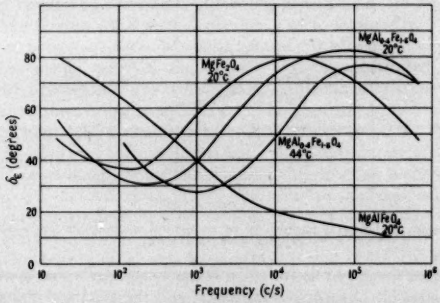


Fig. 8. Dielectric loss versus frequency characteristics of compound magnesium ferrite-aluminates (Fairweather and Frost 1951).

permittivity, and piezoelectricity of the high dielectric loss in a polycrystalline material). David and Welch (1951) have, however, found 'extra' reflections in the x-ray patterns for γ-Fe_2O_3 and for Li ferrite ($Li\ Fe_5O_8$), suggesting some degree of superlattice ordering, but it is not known whether these materials exhibit exceptional dielectric properties. Slowly cooled Cu ferrite (Weil, Bertaut and Bochirol 1950) has a tetragonal structure, but its permittivity at 1 Mc/s is actually lower than that of the quenched cubic form.

In 1938 Blechschmidt observed the variation of dielectric constant and dielectric loss of certain Mn ferrites over a range of radio frequencies, and attempted to simulate their behaviour with artificial dielectrics consisting of graphite–paraffin wax mixtures. Little information has, however, been generally published on the dielectric properties of ferrites, but recent measurements (Fairweather and Frost 1951, Koops 1951), have shown that they are qualitatively similar to those of other polycrystalline semiconductors (see figs. 8 and 9).

The frequency corresponding to the intermediate maximum in the dielectric loss angle δ_ε increases with temperature and varies systematically with the composition, sintering and heat treatment of the material.

Billig and Plessner (1951) have proposed to explain similar frequency characteristics (in another connection) in terms of a particular model of an inhomogeneous dielectric, consisting of higher conductivity dielectric particles separated by lower conductivity dielectric boundary layers. This model is not, however, unique and further evidence is required before a more complete theory can be established.

Ballistic measurements of the permittivity of various ferrites (Fairweather and Frost 1951) have given results ranging from slightly larger than, up to 100

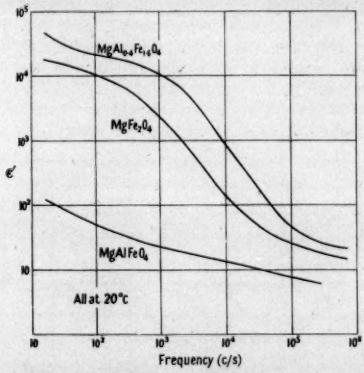


Fig. 9. Dispersion of dielectric permitivity of compound magnesium ferrite-aluminates (Fairweather and Frost 1951).

times, the values measured at 100 c/s. It has been found that some ferrites are more or less strongly absorptive dielectrics, and the electrical absorption shown to be correlated with the presence of H_2O in adsorbed or possibly combined form in the material. Although Pauthenet and Bochirol (1951) have concluded from magnetic saturation evidence that Mg^{2+} ions in Mg ferrite can migrate from B to A sites in the lattice with a rather low activation energy (their value of the temperature θ in eqn. (15) is only 1 220° K), no evidence has been found from electrolysis experiments at room temperature that these ions contribute more than 0·1% to the conductivity. Nevertheless, prolonged washing of certain samples in conductivity water has reduced the absorption considerably, suggesting the presence of some soluble salts.

(iii) *Semiconducting Properties*

According to the electronic energy band of solids, the oxides of the transition metals should be semi-metallic conductors rather than insulators or ordinary semiconductors, since the metallic 3d band should in all cases be only partly occupied by electrons. Yet experience shows that all except Fe_3O_4 are practically insulators when of accurately stoichiometric composition. Slater (1951) has recently proposed to explain the anomaly on the basis that the 'superexchange'

coupling that is present in such oxides also has the effect of splitting the 3d band in such a way that the lower part is completely filled and the upper part completely empty, with a sufficient intervening gap to make the material no more than a weak intrinsic semiconductor. In the exceptional case of Fe_3O_4 the presence of Fe^{3+} and Fe^{2+} ions in neighbouring B sites, with only a small energy barrier to the interchange of an electron between them (equivalent to an interchange of Fe^{2+} for Fe^{3+} ions and vice versa), leads to a relatively high electronic conductivity independently of the usual semiconduction mechanism. Whether Zener's (1951 a, b) mechanism of 'double exchange' is important in this case is not yet clear.

For a simple semiconductor we expect the electronic conductivity σ to vary with the absolute temperature T according to the law

$$\sigma = A \exp(-B/kT), \qquad \ldots\ldots(24)$$

where A and B are constants of the material and k is Boltzmann's constant. In general it is found for ferrites, as for most other semiconductors, that the plot of $\log\sigma$ versus $1/T$ is more or less curved, but approximate values of B may be assigned over particular temperature ranges. For Fe_3O_4 in the neighbourhood of room temperature B is equivalent to about 0·1 ev, while for the low-conductivity ferrites B may reach 0·5 ev or more. The data of Guillaud and Bertrand (1950) for Mg and Mn ferrites show that B decreases with decrease of T below room temperature, a behaviour also found in certain other semiconductors (Billig 1951). Measurements on Mn–Zn, Ni–Zn and Al-substituted Mg ferrites (Eastwood and Fairweather 1951) have shown substantially constant B values over the temperature range 290–530° K. The thermo-electric power, unlike A and B, has been found nearly independent of composition, at a value of approximately $-200\,\mu$v per ° C over the latter temperature range.

The influence of composition, the grinding and sintering techniques, and the heat treatment on the d.c. conductivity–temperature and conductivity–voltage behaviour of certain Mn ferrites was examined by Schulze and Gremmer (1938). Verwey and Haayman (1941) discussed the electronic conductivity of magnetite over a wide temperature range and Verwey (1947) that of several mixed ferrites, while Snoek (1947–9) has interpreted the general characteristics in terms of oxygen defect and the resultant small proportion of magnetite appearing in most ferrites. Snoek's work has shown how the conductivity of a ferrite can be greatly increased by heating in vacuum at 400–600 °C, and restored to its original low value by heating at the same temperature in oxygen at atmospheric pressure. The mechanism may be regarded as that of reduction of the Fe_2O_3 and the oxidation of the Fe_3O_4, with an equilibrium relation of the form

$$6Fe_2O_3 \rightleftharpoons 4Fe_3O_4 + O_2, \qquad \ldots\ldots(25)$$

depending upon temperature and oxygen pressure and upon the amount of free surface available to the oxygen. A densely sintered sample can absorb much less oxygen in this way than a relatively porous one, and a compromise appears to be necessary in commercial production unless repeated regrinding and resintering in oxygen is to be employed.

This oxygen absorption and release must be primarily a surface phenomenon, since the probability of the large oxygen ions migrating through the lattice is very small at all but the highest temperatures. It thus seems likely that adsorbed oxygen becomes built into the ferrite lattice by the migration outwards of some of the more mobile metal ions, the latter having been held in interstitial sites

if the ferrite was stoichiometric before the loss of oxygen. In general the initial and final composition may be slightly non-stoichiometric, and there may be therefore a small fraction of the normally occupied A and B sites vacant (this fraction being too small to be detectable by x-ray diffraction, magnetic saturation, or density measurements).

Much further work remains to elucidate the oxygen absorption mechanism and to correlate it with possible H_2O absorption (exemplified in the γ-Fe_2O_3 structure and in the anomalous discharge current in other ferrites, as already noted).

§ 10. CONCLUDING REMARKS

It has been necessary to curtail the discussion in several of the preceding sections in order to keep the material within the proper bounds of a review of progress. This has been especially the case in connection with magnetic and dielectric dispersion, where the generally published work is as yet rather limited although the significance of the subject, both theoretical and practical, is undoubtedly great.

Attention may be directed at this point to previous work of a review or an original nature, which has been drawn on extensively in compiling the present review. Snoek's book (1947–9) and Néel's paper on ferrimagnetism (1948) are primary sources, to be supplemented by the equally fundamental work of Verwey and his collaborators (1935 onwards).

Session IV of the November 1949 London Symposium on Ferromagnetic Materials arranged by the Institution of Electrical Engineers, and the July 1950 Grenoble Conference on Ferromagnetism and Antiferromagnetism, together provided numerous important papers which have been referred to individually in the preceding text and in the bibliography at the end of this review. Two short reviews have been given by Brockman (1949, 1951), and others by Flegler (1950) and by Harvey, Hegyi and Leverenz (1950). Attention should also be drawn to the work of Shull and his collaborators (1949, 1951) on the application of the neutron diffraction technique. Although beset with difficulties due among other things to the generally large non-magnetic scattering in most ferrites, this technique has already demonstrated the correctness of Néel's theory of the ferrimagnetic structure of magnetite, and promises to be a valuable tool for resolving the structure of other ferrites.

No attempt has been made to give a complete bibliography. Further references will be found in the papers noted.

APPENDIX

NÉEL'S THEORY OF FERRIMAGNETISM

The equivalent Weiss field acting on an A ion (i.e. a Y ion in a A site) is assumed to have the form:

$$h_a = n(\alpha\lambda \mathbf{I}_a + \epsilon\mu \mathbf{I}_b), \qquad \ldots\ldots(1)$$

and similarly for a B ion (i.e. a Y ion in a B site):

$$h_b = n(\beta\mu \mathbf{I}_b + \epsilon\lambda \mathbf{I}_a), \qquad \ldots\ldots(2)$$

where $\mathbf{I}_a$ = magnetization per gram-ion due to A ions alone, $\mathbf{I}_b$ = magnetization per gram-ion due to B ions alone, n is a positive constant, λ = fraction of Y ions in A sites, μ = fraction of Y ions in B sites (so that we always have $\lambda+\mu=1$),

α, β = constants, of positive sign for ferromagnetic, and negative sign for antiferromagnetic, coupling *within* the corresponding A or B group, $\epsilon = +1$ for ferromagnetic and $\epsilon = -1$ for antiferromagnetic coupling *between* the A and B groups of ions.

On account of the interactions between neighbouring magnetic ions and of their different environments, the gram-ion magnetizations I_a and I_b are in general different from each other and from the value inferred for the Fe^{3+} ion from susceptibility–temperature measurements on magnetically diluted paramagnetic salts. The average gram-ion magnetization is

$$\mathbf{I} = \lambda \mathbf{I}_a + \mu \mathbf{I}_b, \qquad \ldots\ldots(3)$$

where vector addition is assumed. When an external field H is applied the resultant magnetization is assumed to follow a Curie law for the A ions and the B ions separately:

$$I_a = C(H + h_a)/T, \qquad \ldots\ldots(4) \qquad I_b = C(H + h_b)/T \qquad \ldots\ldots(5)$$

where C is the Curie constant defined in the usual way for simple paramagnetic ions in weak field at not too low temperatures, and T is the absolute temperature.

In the temperature range in which spontaneous magnetization does not occur, we may solve eqns. (1)–(5) to obtain the resultant paramagnetic susceptibility in terms of the parameters of the material. When $\epsilon = -1$ we find

$$\frac{1}{\chi} = \frac{H}{I} = \frac{T}{C} + \frac{1}{\chi_0} - \frac{\sigma}{T - \theta}, \qquad \ldots\ldots(6)$$

where

$$\frac{1}{\chi_0} = n(2\lambda\mu - \lambda^2\alpha - \mu^2\beta), \qquad \ldots\ldots(7)$$

$$\sigma = n^2 C\lambda\mu[\lambda(1+\alpha) - \mu(1+\beta)]^2, \qquad \ldots\ldots(8)$$

$$\theta = nC\lambda\mu(2 + \alpha + \beta). \qquad \ldots\ldots(9)$$

The last term in eqn. (6) causes a departure from the classical linear Curie–Weiss law at the lower temperatures, and leads to a hyperbolic curvature concave towards the T axis. The curve meets the T axis at the paramagnetic Curie temperature θ_p given by:

$$\theta_p = \tfrac{1}{2} nC[\lambda\alpha + \mu\beta + \{(\lambda\alpha - \mu\beta)^2 + 4\lambda\mu\}^{1/2}]. \qquad \ldots\ldots(10)$$

If θ_p is positive, spontaneous magnetization is to be expected at temperatures below θ_p; otherwise the ferrite will remain paramagnetic at all temperatures. The case of positive θ_p thus leads to a type of ferromagnetism even though the inter-group interaction, represented by $\epsilon = -1$, favours mutually anti-parallel alignment of the group magnetic moments. This type of ferromagnetism Néel has called 'ferrimagnetism'.

Equation (10) applies also for the case of $\epsilon = +1$, and eqn. (6) has the same form, but χ_0, σ and θ in (6) are modified somewhat. The same general behaviour, though with less possible variety, is to be expected in this case.

Considering next the temperature range in which spontaneous magnetization occurs, the magnitudes of the spontaneous gram-ion magnetization $\mathbf{I}_{as}$ and $\mathbf{I}_{bs}$ are to be found such that the following equations are solved simultaneously:

$$\mathbf{I}_{as} = M B_j[M h_a/RT], \qquad \ldots\ldots(11) \qquad \mathbf{I}_{bs} = M B_j[M h_b/RT], \qquad \ldots\ldots(12)$$

where M = absolute saturation magnetization per gram-ion of the Y ions, R = universal gas constant, and $B_j[\,] = $ Brillouin function, as employed in the modified Weiss theory of the magnetization of ferromagnetic metals (cf. Seitz 1940). The values of h_a and h_b to be inserted in eqns. (11) and (12) are given by

(1) and (2), and the magnitude of the overall resultant spontaneous magnetization becomes:

$$I_s = |\lambda I_{as} + \epsilon\mu I_{bs}|. \qquad \ldots\ldots(13)$$

Néel has shown (1948) that eight different types of temperature variation of I_s and $1/\chi$ are possible on the basis of the preceding discussion. These are illustrated by the curves of fig. 3. The labelling of the curves identifies corresponding areas in a diagram representing the α, β plane divided up in a manner dependent upon the ratio λ/μ and upon the sign of ϵ. Antiferromagnetism arises naturally in Néel's theory as the special case of ferrimagnetism in which $\epsilon = -1$, $\lambda = \mu$ and $\alpha = \beta$. This has been discussed in a recent paper by Street (1951). The reader is referred to the original paper (Néel 1948) for the rather elaborate explanation.

REFERENCES

ADAM, G. D., and STANDLEY, K. M., 1950, *Proc. Phys. Soc.* B, **63**, 820.

AHMED, M. S., 1950, *Nature, Lond.*, **165**, 246.

AKHIESER, A., 1946, *J. Phys., U.S.S.R.*, **10**, 137.

AKHIESER, A., and POMERANCHUK, I., 1944, *J. Phys., U.S.S.R.*, **8**, 206.

ANDERSON, P. W., 1950, *Phys. Rev.*, **79**, 350.

ANDREWS, W. K., 1951, *Trans. Brit. Ceram. Soc.*, **50**, 47.

VAN ARKEL, A. E., VERWEY, E. J. W., and VAN BRUGGEN, M. G., 1936, *Rec. Trav. Chim. Pays-Bas*, **55**, 331.

BARBLAN, F., BRANDENBURGER, E., and NIGGLI, P., 1944, *Helv. Chim. Acta*, **27**, 88.

BARTH, T. F. W., and POSNJAK, E., 1932, *Z. Kristallogr.*, **82**, 325.

BECKER, R., 1938, *Phys. Z.*, **39**, 856; 1939, *Ferromagnetismus* (Berlin: Springer), pp. 237–42; 1951, *J. Phys. Radium*, **12**, 332.

BELJERS, H. G., and POLDER, D., 1950, *Nature, Lond.*, **165**, 800.

BERGER, S. V., 1948, *Festskrift tillägnad J. Arvid Hedvall* (Göteborg).

BERTAUT, E. F., 1951, *J. Phys. Radium*, **12**, 252.

BICKFORD, L. R., 1949 a, *Phys. Rev.*, **75**, 1298 (A); 1949 b, *Ibid.*, **76**, 137; 1950, *Ibid.*, **78**, 449.

BILLIG, E., 1951, *Proc. Phys. Soc.* A, **64**, 752.

BILLIG, E., and PLESSNER, K. W., 1951, *Proc. Phys. Soc.* B, **64**, 361.

BIRKS, J. B., 1947, *Nature, Lond.*, **160**, 535; 1948 a, *Proc. Phys. Soc.*, **60**, 282; 1948 b, *Phys. Rev.*, **74**, 843; 1948 c, *Ibid.*, **74**, 988; 1950, *Proc. Phys. Soc.* B, **63**, 65.

BIZETTE, H., 1951, *J. Phys. Radium*, **12**, 161.

BLECHSCHMIDT, E., 1938, *Phys. Z.*, **39**, 212.

BLOEMBERGEN, N., 1950, *Phys. Rev.*, **78**, 572.

BRATTAIN, W. H., and BRIGGS, H. B., 1949, *Phys. Rev.*, **75**, 1705.

BROCKMAN, F. G., 1949, *Elect. Engng.*, **68**, 1077; 1951, *Ibid.*, **70**, 489.

BROCKMAN, F. G., DOWLING, P. H., and STENECK, W. G., 1949, *Phys. Rev.*, **75**, 1298 (A), 1440; 1950, *Ibid.*, **77**, 85.

CHARLESWORTH, G., and LONG, F. A., 1938–9, *Proc. Leeds Phil. Lit. Soc.*, **3**, 515.

CHEVALLIER, R., 1951, *J. Phys. Radium*, **12**, 172.

CHEVALLIER, R., and GIRARD, J., 1950, *Bull. Soc. Chim. Fr.*, **17**, 576.

CHING HSIEN LI, 1932, *Phys. Rev.*, **40**, 1002.

DAVID, I., and WELCH, A. J. E., 1951, Paper for 12th Internat. Congress on Pure and Applied Chemistry, New York.

DOMENICALI, C. A., 1950, *Phys. Rev.*, **78**, 458.

DÖRING, W., 1948, *Z. Naturforsch.*, **3a**, 373.

DOWLING, P. H., and BROCKMAN, F. G., 1949, *Phys. Rev.*, **75**, 1298.

EASTWOOD, G. W., and FAIRWEATHER, A., 1951, *Post Office Engineering Dept. Research Report.*

FAIRWEATHER, A., and FROST, E. J., 1951, *Post Office Engineering Dept. Research Report.*

FALLOT, M., and MARONI, P., 1951, *J. Phys. Radium*, **12**, 256.

FLEGLER, E., 1950, *Arch. Elektrotech.*, **40**, 4.

FOËX, G., 1951, *J. Phys. Radium*, **12**, 153.

FOËX, G., and SERRES, A., 1950, *C. R. Acad. Sci., Paris*, **230**, 729.

FORESTIER, H., 1931, *C. R. Acad. Sci., Paris*, **192**, 842.

FORRER, R., 1951, *J. Phys. Radium*, **12**, 402.
FYFE, W. S., 1949, *Nature, Lond.*, **164**, 790.
GALT, J. K., 1951, *Phys. Rev.*, **83**, 208.
GALT, J. K., MATTHIAS, B. T., and REMEIKA, J. P., 1950, *Phys. Rev.*, **79**, 391.
GALT, J. K., YAGER, W. A., REMEIKA, J. P., and MERRITT, F. R., 1951, *Phys. Rev.*, **81**, 470.
GORTER, E. W., 1950 a, *C. R. Acad. Sci., Paris*, **230**, 192; 1950 b, *Nature, Lond.*, **165**, 798.
GREENWALD, S., and SMART, J. S., 1950, *Nature, Lond.*, **166**, 523.
GUILLAUD, C., 1950, *J. Rech. Cent. Nat. Rech. Sci.*, **12**, 113; 1951 a, *J. Phys. Radium*, **12**, 239; 1951 b, *C. R. Acad. Sci., Paris*, **232**, 944.
GUILLAUD, C., and BERTRAND, R., 1950, *J. Recherches, C.N.R.S.*, No. 11, 73.
GUILLAUD C., and MICHEL A., 1951, *J. Phys. Radium*, **12**, 65.
GUILLAUD, C., and ROUX, M., 1949, *C. R. Acad. Sci., Paris*, **229**, 1133.
GUILLAUD, C., and SAGE, M., 1950, *J. Phys. Radium*, **11**, 4E.
HÄGG, G., 1935, *Z. phys. Chem.* B, **29**, 95.
HÄGG, G., and SODERHOLM, G., 1935, *Z. phys. Chem.* B, **29**, 88.
HARVEY, R. L., HEGYI, I. J., and LEVERENZ, H. W., 1950, *R. C. A. Rev.*, **11**, 321.
HERRING, C., and KITTEL, C., 1951, *Phys. Rev.*, **81**, 869.
HEWITT, W. H., 1948, *Phys. Rev.*, **73**, 1118; *Ibid.*, **74**, 1211(A).
HILPERT, G., 1909, German Patent Nos. 226347 and 227787; *Ber. dtsch. chem. Ges.*, **42**, 2248.
VON HIPPEL, A., 1950, *Rev. Mod. Phys.*, **22**, 236.
VON HIPPEL, A., BRECKENRIDGE, R. G., CHESLEY, F. G., and TISZA, L., 1946, *Industr. Engng. Chem.* **38**, 1097.
HOFFMANN, A., 1938, *Naturwissenschaften*, **26**, 431.
HUZIMURA, T., 1950, *J. Phys. Soc. Japan*, **5**, 56.
INTERNATIONAL CRITICAL TABLES, 1926, (New York : McGraw-Hill Co.), **6**.
ISKENDERIAN, H. P., 1949, *Phys. Rev.*, **76**, 175 (A).
JONES, G. O., and ROBERTS, F. F., 1951, *Post Office Engineering Dept. Research Report.*
JONKER, G. H., and VAN SANTEN, J. H., 1950, *Physica*, **16**, 337.
KATAYAMA, T., 1948, *J. Phys. Soc. Japan*, **4**, 202.
KAY, H. F., and VOUSDEN, P., 1949, *Phil. Mag.*, **40**, 1019.
KINGSNORTH, S. W., 1950, *Studies in the Chemistry and Structure of Magnesium Ferrite*, Ph.D. Thesis, Imperial College, London University.
KINGSNORTH, S. W., and WELCH, A. J. E., 1951, Paper for 12th Internat. Congress on Pure and Applied Chemistry, New York.
KITTEL, C., 1946, *Phys. Rev.*, **70**, 281; 1948, *Ibid.*, **73**, 155; 1949, *Phys. Rev.*, **76**, 176(A), 743; 1951 a, *J. Phys. Radium*, **12**, 291; 1951 b, *Phys. Rev.*, **82**, 565.
KOOPS, C. G., 1951, *Phys. Rev.*, **83**, 121.
KORDES, E., and ROTTIG, 1951, *Z. anorg. Chem.*, **264**, 34.
KORNETZKI, M., 1951, *Z. angew. Phys.*, **3**, 5.
KRAMERS, H. A., 1934, *Physica*, **1**, 182.
KUSSMAN, A., and NITKA, H., 1938, *Phys. Z.*, **39**, 208.
LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Sowjet.*, **8**, 153.
LATIMER, K. E., and MACDONALD, H. B., 1950, *Proc. Instn. Elect. Engrs.*, **97**, Pt. II, 257.
LEGG, V. E., 1936, *Bell Syst. Tech. J.*, **15**, 39.
MAXWELL, L. R., SMART, J. S., and BRUNAUER, S., 1949, *Phys. Rev.*, **76**, 459 (A).
MICHEL, A., CHAUDRON, G., and BENARD, J., 1951, *J. Phys. Radium*, **12**, 189.
MILLAR, R. W., 1929, *J. Amer. Chem. Soc.*, **51**, 215.
NÉEL, L., 1944, *Cah. de Phys.*, **25**, 21.
NÉEL, L., 1947, *Ann. Univ. Grenoble*, **22**, 299.
NÉEL, L., 1948, *Ann. Phys., Paris*, **3**, 1937; 1950 a, *C. R. Acad. Sci., Paris*, **230**, 190; 1950 b, *Ibid.*, **230**, 375; 1950 c, *J. Phys. Radium*, **11**, 49; 1950 d, *Physica*, **16**, 350; 1951 a, *J. Phys. Radium*, **12**, 258; 1951 b, *Ibid.*, **12**, 339.
NÉEL, L., and BROCHET, P., 1950, *C. R. Acad. Sci., Paris*, **230**, 280.
NICKS, P. F., and WELCH, A. J. E., 1951, Paper for 12th Internat. Congress on Pure and Applied Chemistry, New York.
OKAMURA, T., 1951, *Nature, Lond*, **168**, 162.
OKAMURA, T., TORIZUKA, Y., and KOJIMA, Y., 1951, *Phys. Rev.*, **82**, 285.
PAUTHENET, R., and BOCHIROL, L., 1951, *J. Phys. Radium*, **12**, 249.
POLDER, D., 1950, *Proc. Instn. Elect. Engrs.*, **97**, Pt. II, 246; 1949, *Phil. Mag.*, **40**, 99.
POSNJAK, E., 1930, *Amer. J. Sci.*, **19**, 67.

POULIS, N. J., VAN DEN HANDEL, J., UBBINK, J., POULIS, J. A., and GORTER, C. J., 1951, *Phys. Rev.*, **82**, 552.
QUITTNER, V., 1909, *Ann. Phys., Lpz.*, **30**, 289.
RADO, G. T., 1950, *Advances in Electronics*, Vol. II, 251 (New York : Academic Press) ; 1951, *Phys. Rev.*, **83**, 821.
RADO, G. T., and TERRIS, A., 1951, *Phys. Rev.*, **83**, 177.
RADO, G. T., WRIGHT, R. W., and EMERSON, W. H., 1950, *Phys. Rev.*, **80**, 273
RICE, O. K., 1940, *Electronic Structure and Chemical Binding* (New York : McGraw-Hill).
RICHARDS, C. E., BUCKLEY, S. E., BARDELL, P. R., and LYNCH, A. C., 1950, *Proc. Instn. Elect. Engrs.*, **97**, Pt. II, 236.
RICHTER, G., 1937, *Ann. Phys., Lpz.*, **29**, 605.
ROBERTS, F. F., 1951 a, *J. Phys., Radium*, **12**, 305 ; 1951 b, *Ibid.*, **12**, 301.
ROBERTS, H. S., and MERWIN, H. E., 1931, *Amer. J. Sci.*, **21**, 145.
ROOKSBY, H. P., and TOMBS, N. C., 1951, *Nature, Lond.*, **167**, 364.
RUDERMAN, I. W., 1949, *Phys. Rev.*, **76**, 1572.
SAGE, M., and GUILLAUD, C., 1950, *C. R. Acad. Sci., Paris*, **230**, 1751.
VAN SANTEN, J. H., and JONKER, G. H., 1951, *J. Phys. Radium*, **12**, 202.
SCHULZE, A., and GREMMER, W., 1938, *Phys. Z.*, **39**, 205.
SEITZ, F., 1940, *Modern Theory of Solids* (New York : McGraw-Hill).
SHÔJI, H., 1935, *Sci. Rep. Tôhoku Univ.*, **24**, 250.
SHULL, C. G., 1951, *Phys. Rev.*, **81**, 626.
SHULL, C. G., and SMART, J. S., 1949, *Phys. Rev.*, **76**, 1256.
SHULL, C. G., STRAUSER, W. A., and WOLLAN, E. O., 1951, *Phys. Rev.*, **83**, 333.
SHULL, C. G., WOLLAN, E. O., and STRAUSER, W. A., 1951, *Phys. Rev.*, **81**, 483.
SLATER, J. C., 1936, *Phys. Rev.*, **49**, 537, 931 ; 1950, *Ibid.*, **78**, 748 ; 1951, *Ibid.*, **82**, 538.
SNOEK, J. L., 1936, *Physica*, **3**, 463 ; 1946, *Philips Tech. Rev.*, **8**, 353 ; 1947, *Nature, Lond.*, **160**, 90 ; 1947–9, *New Developments in Ferromagnetic Materials* (Amsterdam : Elsevier) ; 1948, *Physica*, **14**, 207 ; 1950, *Ibid.*, **16**, 333 ; 1951, *J. Phys. Radium*, **12**, 228.
STONER, E. C., 1948, *Rep. Prog. Phys.*, **11**, 43 (London : Physical Society); 1950, *Ibid.*, **13**, 83.
STREET, R., 1951, *Sci. Progr.*, **39**, 358.
STREET, R., and WOOLLEY, J. C., 1949, *Proc. Phys. Soc.* A, **62**, 743.
STRUTT, M. J. O., 1950, *Wireless Engr.*, **27**, 277.
TELLEGEN, B. D. H., 1948, *Philips Res. Rep.*, **3**, 81.
TOMBS, N. C., and ROOKSBY, N. P., 1950, *Nature, Lond* **165**, 442.
TROUNSON, E. P., BLEIL, D. F., WANGSNESS, R. K., and MAXWELL, L. R., 1950, *Phys. Rev.*, **79**, 542.
VAN VLECK, J. H., 1951 a, *J. Phys. Radium*, **12**, 262 ; 1951 b, *Physica*, **17**, 234.
VERWEY, E. J. W., 1935, *Z. Kristallogr.*, **91**, 65 ; 1947, *Philips Tech. Rev.*, **9**, 46 ; 1947, *Chem. Weekbl.*, **43**, 229.
VERWEY, E. J. W., and DE BOER, J. H., 1936, *Rec. Trav. Chim. Pays-Bas*, **55**, 531.
VERWEY, E. J. W., DE BOER, J. H., and VAN SANTEN, J. H., 1948, *J. Chem. Phys.*, **16**, 1091.
VERWEY, E. J. W., and HAAYMAN, P. W., 1941, *Physica*, **9**, 979.
VERWEY, E. J. W., HAAYMAN, P. W., and HEILMANN, E. L., 1947, *Chem. Weekbl.*, **9**, 185.
VERWEY, E. J. W., HAAYMAN, P. W., and ROMEIJN, F. C., 1947, *J. Chem. Phys.*, **15**, 181.
VERWEY, E. J. W., and HEILMANN, E. L., 1947, *J. Chem. Phys.*, **15**, 174.
WANNIER, G. H., 1950, *Phys. Rev.*, **79**, 357.
WEIL, L., 1951, *J. Phys. Radium*, **12**, 260.
WEIL, L., BERTAUT, F., and BOCHIROL, L., 1950, *J. Phys., Radium*, **11**, 208.
WEIS, A., 1951, *Elektrotech. Nachr.*, **5**, 213.
WEISS, P., 1896, *J. Phys. Radium*, **8**, 435.
WEISS, P., and FORRER, R., 1929, *Ann. Phys., Paris*, **12**, 297.
WELCH, A. J. E., NICKS, P. F., FAIRWEATHER, A., and ROBERTS, F. F., 1950, *Phys. Rev.*, **77**, 151.
WENT, J. J., and WIJN, H. P. J., 1951, *Phys. Rev.*, **82**, 269.
WIEBERDINK, A., and KRONIG, R., 1948, *Nature, Lond.*, **162**, 527.
WOLOGDINE, S., 1909, *C. R. Acad. Sci., Paris*, **148**, 776.
WYCKOFF, R. W. G., 1930, 1934, 1935, *Structure of Crystals* (New York : Chem. Catalog Co.).
YAGER, W. A., MERRITT, F. R., and GUILLAUD, C., 1951, *Phys. Rev.*, **81**, 477.
ZENER, C., 1951a, *Phys. Rev.*, **81**, 440; 1951 b, *Ibid.*, **82**, 403.

THE BAND STRUCTURE OF METALS

By G. V. RAYNOR*
Department of Metallurgy, University of Birmingham

CONTENTS

	PAGE
§ 1. Introduction	173
§ 2. The alkali metals	182
§ 3. Metals of group IB of the Periodic Table	193
§ 4. The divalent metals	199
§ 5. Aluminium	213
§ 6. The transitional metals	221
§ 7. The band structures of other metals and non-metals	227
§ 8. The electronic structure of alloys	232
§ 9. Conclusion	243
References	246

ABSTRACT. The development of the electron theory of metals is briefly reviewed, and the electronic structures of a selection of monovalent and polyvalent metals are surveyed. Sodium and copper are discussed in detail as examples of the alkali metals and the group IB metals respectively. Agreement between experimental and theoretical values of the binding energies, the lattice constants, the compressibilities and the elastic constants is adequate, but difficulties arise in connection with the calculation of the electronic interactions. These difficulties are increased when divalent metals are considered; the divalent metals calcium, magnesium and beryllium are discussed in this connection. The electronic structures of these metals involve overlap of electrons from the first Brillouin zone, and for magnesium alloys, lattice spacing relationships may be satisfactorily interpreted on this basis. For trivalent aluminium, the mathematical difficulties are more serious; it has, however, been possible to explain the elastic anisotropy of this metal, and the calculated variation of elastic constants with electron/atom ratio enables phenomena encountered in aluminium-rich alloys to be interpreted. The nature of transitional metals is discussed in detail. The possibility of the extension of the theory to alloys is reviewed, with reference to alloys based on copper and on nickel.

§ 1. INTRODUCTION

THE problem of the fundamental nature of the metallic state is of long standing, and important contributions have been made not only by physicists, but also by many chemists and metallurgists. It is, however, to the pure physicists that the major advances are due, and it is the more physical aspects of the structure of metals which will be considered below.

The most characteristic property of a metal is its ability to conduct electricity, and it was natural that the earlier attempts to work out a theory of metals should concentrate on providing a reasonable mechanism for conduction. It was early appreciated that the carrier of the electric current in a metal must be the electron, and direct experimental evidence for this was provided by the work of Tolman and Stewart (1916, 1917)†. These workers investigated the current pulse observed when a metallic conductor, in rapid rotational motion, was suddenly stopped; inertia effects were observed which could be attributed to current carriers having

* Temporarily at the Institute for the Study of Metals, University of Chicago, Chicago 37, Illinois.

† See also Tolman, Karrer and Guernsey (1923) and Tolman and Mott-Smith (1926).

a charge/mass ratio closely similar to that of the electron. The converse experiment, the acceleration of a metallic conductor owing to the inertia of the electric current, was suggested by Maxwell, and carried out by Barnett (1931), who also obtained a charge/mass ratio for the carrier consistent with that of the electron. In recent years, a much more precise determination of the charge/mass ratio has been made by Kettering and Scott (1944). Cylindrical coils of copper or aluminium wire were supported in such a way as to form a torsional pendulum, and the amplitude of vibration with a certain current flowing was measured. On reversing the current, the amplitude was changed, owing to the angular momentum of the current, and precise observation of this change enabled the value of charge/mass to be determined. It may be shown that the amplitude change $\Delta\theta$ is related to the mass and charge of the current carrier by the relation

$$\Delta\theta = \left(\frac{2iANP}{\pi I}\right)\left(\frac{m}{e}\right), \qquad \ldots\ldots(1)$$

where i is the current, m the mass of the current carrier and e its charge, P the period of the oscillation and I the moment of inertia of the coil ; A and N are constants of the apparatus. The average value for m/e obtained from several experiments carried out with various values of i was $5{\cdot}69\times10^{-9}$ g/coulomb, as compared with the accepted value of $5{\cdot}68\times10^{-9}$ g/coulomb for the electron itself. Our picture of a metal, therefore, must be that it contains electrons which are free to move, and which may, when under the influence of an electric field, carry a current through the material. The theory of metals, in fact, must be an electronic theory.

The free atom of sodium contains one 3s electron which is relatively loosely bound and may be removed in chemical reactions, leaving behind the sodium ion, consisting of a nucleus and closed shells of electrons, bearing one positive charge. When a large number of sodium ions come together to form the body-centred cubic arrangement of solid sodium, we must suppose that the lattice points are occupied by the ions, while the 3s electrons in some manner become free enough to move throughout the lattice. The first electronic theory of metals along these lines was due to Lorentz (1916), who postulated that the spaces between the ions which occupied the lattice sites were filled by electrons which were essentially free, and subject only to external applied electric fields and forces arising from elastic collisions with the ions, which were considered to act like hard spheres. The application of an electric field then accelerates the electrons, producing a drift of charges and hence an electric current. Linear increase of current with time is prevented by collisions of electrons with ions, which are equivalent to a resisting force proportional to the electronic velocity. Under the influence of an external field E the electron is subject to a resultant force of $-eE-kv$, so that $mdv/dt+kv=-eE$. The solution of this equation is the sum of two parts, namely, the steady drift $v=-eE/k$ and the transient velocity $v=ge^{-kt/m}$, where g is a constant, which is damped so rapidly that it is entirely negligible under normal conditions. Thus the current density, which, for N electrons per unit volume, is $-Nev$, may be written $(Ne^2/k)E$, where Ne^2/k is the specific conductivity of the metal. This simple model, therefore, accounts satisfactorily for Ohm's law, but it may easily be seen that it is otherwise unacceptable. The theory was based essentially upon the application of classical mechanics, so that, on applying heat to the metal, the electrons must be assumed, in equilibrium, to take up the kinetic energy distribution characteristic of gas molecules at the same temperature. The

electrons would thus be expected to contribute appreciably to the specific heat of the metal, whereas it is well known that at ordinary temperatures their electronic specific heat is very small indeed. A further difficulty in the theory lay in the interpretation of the vanishingly small resistance of metals as the absolute zero of temperature is approached, which made it necessary to assume that the electronic mean free path became infinite at absolute zero. Although, therefore, the Lorentz theory allowed a preliminary picture of the nature of a metal to be built up, it was seriously incomplete.

The theory was considerably advanced by Sommerfeld, Houston and Eckart (1928), who applied certain aspects of the quantum theory to the preliminary model developed by Lorentz. This development was of such importance that it may be considered in some detail. It will be recalled that difficulties associated with the application of the Bohr theory of the atom, and in particular with the arbitrary introduction of quantum restrictions to the angular momentum of an electron rotating in an orbit otherwise interpreted according to classical mechanics, had led to the development of the methods of wave mechanics. According to the wave-mechanical treatment, the energies which the valency electron of hydrogen, for example, could possess are determined by the solutions of the Schroedinger equation:

$$\nabla^2\psi + \frac{2m}{\hbar^2}[E - V(r)]\psi = 0, \qquad \ldots\ldots(2)$$

where E is the total energy of the electron, m its mass, and $\hbar$ is equal to $h/2\pi$ where h is Planck's constant; $V(r)$ is the potential energy of the electron at a distance r from the nucleus, and, for the hydrogen atom, is equal to $-e^2/r$. It is only possible to obtain acceptable solutions for the function ψ for certain definite values of E, which correspond to the energies of the successive stationary states considered in the Bohr theory. The essential point is that quantization follows logically from the use of the wave-mechanical method, and does not have to be introduced arbitrarily. The function ψ is known as the wave function, and its meaning in the physical sense may be illustrated by reference to the hydrogen atom. The acceptable solutions are bounded, tending to zero at infinity; when the atom is in a stationary state defined by such a solution, $|\psi(r)|^2 d\tau$ is equal to the probability that the valency electron is in the volume element $d\tau$ at the point r, or alternatively, $-e|\psi(r)|^2$ is the average charge density due to the valency electron in the atom at the point r.

When multi-atomic systems are considered, two main methods of approach to the problem of calculating the energies corresponding to various states of the system may be used. According to the Heitler–London scheme, it is assumed that the wave functions are large only about single atoms or ions. The type of solution obtained is accurate only when the atoms of the multi-atomic system are relatively widely separated, and in cases where the atomic or ionic properties of the material under discussion are predominant. The method is convenient for insulators such as sodium chloride and other ionic materials, but has not been applied with much success to metals, owing to mathematical difficulties. For metals, a more convenient method was developed, mainly by Bloch (1928), according to which, by contrast, it is assumed that each wave function extends over the entire multi-atomic system, and has equal amplitude at equivalent atoms. On this model, therefore, it is considered that the atoms of the system are affected by aggregation to such an extent that the valency electrons belong to the entire aggregate rather than to a single atom.

The wave-mechanical type of treatment may thus be applied to the case of a metal, though the results are much less definite than for the hydrogen atom because of the difficulty of defining the precise potential field in which a given electron moves. It is possible, however, to consider each valency electron as moving in the field of the positive ions, plus the average field of all the other electrons, which may be obtained by various approximational methods. If the potential energy of an electron in this field is denoted by $V(r)$, then the wave functions $\psi(r)$ for the electron are solutions of Schroedinger's equation as written above (equation (2)). It is easy to appreciate that the potential field within a metal is periodic with the period of the lattice. Near each ion there is a deep potential trough; between each ion, however, the potential rises to a rounded peak, while at the surface of the metal there is a sharp potential rise. The potential energy of an electron in a metal will therefore have the form of fig. 1 along a line passing

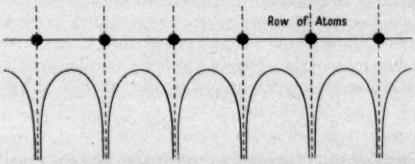


Fig. 1. Form of periodic field in a metal (line passing through nuclei).

through nuclei. It is, however, difficult to carry out calculations for such a field, and Sommerfeld, as an approximation, considered the field inside the metal to be *uniform*, but rising sharply at the surface. Each valency electron is characterized by a wave function ψ which is a solution of Schroedinger's equation for a particle moving in a constant potential field V_e. We may, for simplicity, assume V_e to be zero, and consider a 'one-dimensional' metal (i.e. motion of the electrons is confined to one direction). The appropriate wave functions are solutions of the Schroedinger equation:

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2} E\psi = 0 \qquad \ldots\ldots(3)$$

subject to the condition that the wave functions must vanish at the boundaries of the metal. If the length of the 'one-dimensional' metal is L, the relevant solutions are:

$$\psi = (2/L)^{1/2} \sin(\pi n x/L) \qquad \ldots\ldots(4)$$

where n is an integer, and the corresponding energy levels for the electrons are given by the relation:

$$2m E_n/\hbar^2 = \pi^2 n^2/L^2. \qquad \ldots\ldots(5)$$

The separation of the energy levels depends, therefore, upon L, and, for ordinary macroscopic dimensions, is extremely small, so that the energy levels may for many purposes be considered as forming a continuum. One of Sommerfeld's main contributions was to apply the Fermi–Dirac statistics to the investigation of the manner in which these available energy levels are occupied by electrons. The Fermi–Dirac statistics are based upon the Pauli exclusion principle, according to which no electron can have exactly the same quantum state as another electron. Thus, in the free atom, each energy level is occupied by two electrons only, which differ in the sign of their spin quantum number, but have the remaining quantum

numbers the same. In the metal, each energy level can similarly accommodate two electrons. Into the lowest level, only two electrons may be inserted; of the remainder, only two may enter the next highest level, and so on until all the available electrons are accommodated. According to the Sommerfeld model, therefore, if the metal contains N electrons, the $N/2$ lowest energy levels are occupied and the remainder vacant. The above treatment may easily be extended to the three-dimensional case, when it is found that, for a cube of side L, the allowable wave functions and energy levels are given by the relations :

$$\psi = \sin\frac{\pi n_1 x}{L}\sin\frac{\pi n_2 y}{L}\sin\frac{\pi n_3 z}{L} \qquad \ldots\ldots(6)$$

$$E = \frac{\hbar^2}{2m}\frac{\pi^2}{L^2}(n_1^2+n_2^2+n_3^2). \qquad \ldots\ldots(7)$$

At absolute zero, therefore, the metallic electrons occupy a whole series of energy levels, such that the most energetic electron has an energy of several electron volts. The reason for the negligible electronic specific heat at ordinary temperatures now becomes understandable. It is only for the electrons at the top of the energy distribution that vacant energy levels exist, to which thermally excited electrons may be promoted; it is thus impossible for the total number of electrons to take part in the process, and the resultant effect is negligible at ordinary temperatures. The relationship between the number of electrons and the maximum possible electronic energy follows from the circumstance that the number of occupied states ($N/2$) must equal the number of sets of integers n_1, n_2 and n_3 for which the above expression $(n_1^2+n_2^2+n_3^2)\hbar^2\pi^2/2mL^2$ is less than $E_{\max}$. From this, it may be shown that

$$E_{\max} = \left(\frac{3}{\pi}\right)^{2/3}\frac{\pi^2\hbar^2}{2m}\left(\frac{N}{\Omega}\right)^{2/3} \quad \text{or} \quad N = \frac{\Omega}{3\pi^2}\left[\frac{2m}{\hbar^2}E_{\max}\right]^{3/2} \qquad \ldots\ldots(8)$$

where Ω is the volume of the specimen. It is also important to define the number of electronic states $N_s(E)\,dE$ which correspond to energies between E and $E+dE$ in a volume Ω. Since N_s (the number of states, or $N/2$) is equal to

$$\frac{\Omega}{6\pi^2}\left[\frac{2m}{\hbar^2}E\right]^{3/2}, \qquad \ldots\ldots(9)$$

$N_s(E)\,dE$ may be seen to be equal to

$$\frac{\Omega}{4\pi^2}\left[\frac{2m}{\hbar^2}\right]^{3/2}E^{1/2}\,dE. \qquad \ldots\ldots(10)$$

The total electronic energy of all the electrons in volume Ω is

$$2\int_0^{E_{\max}} N_s(E)E\,dE \quad \text{or} \quad \frac{2\Omega}{4\pi^2}\left[\frac{2m}{\hbar^2}\right]^{3/2}\frac{2}{5}E_{\max}^{5/2} \quad \text{or} \quad \tfrac{3}{5}NE_{\max}. \qquad \ldots\ldots(11)$$

The average energy of the valency electrons thus equals $\frac{3}{5}E_{\max}$; this quantity is usually denoted E_F, the Fermi energy.

The Sommerfeld model thus enables the low electronic specific heats of metals at ordinary temperatures to be interpreted, and allows the calculation of electronic energy levels subject to its initial assumptions. It is still incomplete, however, since no explanation can be given for the characteristic differences which exist between true metallic conductors, semiconductors and insulators. At first sight, similar conclusions should be reached for all types of solid materials. The reason for the inadequacy is that no account has been taken of the periodic nature of the

potential field. This problem was investigated by Bloch (1928) and later by Brillouin (1931). To describe the special effects introduced by consideration of the periodic field, it will be convenient to consider, not wave functions of the type $\psi = \sin(\pi n/L)$ which represent standing waves, but wave functions of the type $\psi = e^{ikx}$ which represent motion in a given direction. To correspond to this form of wave function we have to consider, not a cube of side L, but a ring of material, in which a current can flow continuously, of circumference L. This length must correspond with an integral number of wavelengths of the appropriate wave functions, which means that k must be defined as $2\pi n/L$. The wavelength of the function is L/n, so that k, the wave number, is $2\pi/\lambda$ in this treatment.

The periodicity of the potential field in which the electrons move prevents the wave functions from having the simple form discussed above, and it is necessary to write $\psi = e^{ikx}u_k(x, y, z)$ where u is a function, which depends on k, and which is periodic with the period of the lattice. This wave function may be regarded as representing a plane wave ($\lambda = 2\pi/k$) travelling in the direction of k, modulated by the periodic lattice field. Again considering one dimension for simplicity, the appropriate Schroedinger equation is

$$\frac{d^2\psi}{dx^2} + \frac{2m}{\hbar^2}(E - V)\psi = 0, \qquad \ldots\ldots(12)$$

where V is periodic with the period of the lattice. It may be shown that this equation gives solutions of two different types:

$$\psi = e^{\pm\mu x}u_k(x) \quad \text{and} \quad \psi = e^{\pm ikx}u_k(x) \qquad \ldots\ldots(13)$$

where, as before, $u_k(x)$ is periodic with the period of the lattice and μ is a constant. Only the solutions of the second type correspond to stationary states of the valency electrons in the lattice. The first type corresponds to ranges of energy for which no permissible electronic state exists. Consideration of the real nature of the field in which the valency electrons move, therefore, has led to the important result that the electrons in a metal do not necessarily occupy a continuous series of energy levels from $E = 0$ (measured from the bottom of the valency electron energy distribution) to $E = E_{max}$, but that between $E = 0$ and $E = E_{max}$ there may exist ranges of forbidden energy, the position of which on the energy scale will depend upon the periodicity of the lattice. The determination of the wave functions is a matter of some difficulty, and the procedure adopted for real metals will be more fully discussed below. An indication of the effect of the periodic field may, however, be obtained by considering a one-dimensional metal in which V, the periodic field, is small compared with the electronic kinetic energy. The wave function for an electron has the form $\psi = e^{ikx}u(x)$, and u, being periodic with the period a of the lattice (a = interatomic distance), may be expanded by Fourier methods. Insertion of the expansion into the appropriate Schroedinger wave equation enables a solution to be obtained, subject to certain simplifying assumptions, from which it is found that the energy E is a continuous function of k, the wave number, from $k = 0$ to $k = \pi/a$. At this value of k there is a discontinuity in E; between $k = \pi/a$ and $2\pi/a$ the function is again continuous, so that the plot of energy against wave number appears as in fig. 2. Electrons cannot possess the energy values lying within the gaps ΔE, so that we reach the conclusion that in a metallic crystal the energies which electrons may possess lie in sharply defined bands, each band separated from that above and below by energy discontinuities. In three dimensions, considering a simple cubic lattice, and a

wave number corresponding to a given direction, it is similarly found that discontinuities in E arise for certain values of $\mathbf{k}$, where $\mathbf{k}$ is a vector quantity. The theory shows that, if the vector $\mathbf{k}$ is represented by the coordinates k_x, k_y, k_z, then:

$$k_x n_1 + k_y n_2 + k_z n_3 = \pi(n_1^2 + n_2^2 + n_3^2)/a. \qquad \ldots\ldots(14)$$

This relation corresponds to $k = n\pi/a$ for the one-dimensional case. These critical values of the wave number are exactly those for which Bragg reflection of the corresponding wave motion takes place from planes of atoms with the Miller indices $(n_1 n_2 n_3)$. This may be appreciated by considering an example. In the simple cubic structure, the planes of lowest indices which can diffract electrons are those of the {100} family. For Bragg reflection, the condition is $\lambda = 2d \cos\theta$, where θ is the angle between the direction of propagation of the wave and the normal to the planes {100}; in the present case $d = a$. Thus, for this direction,

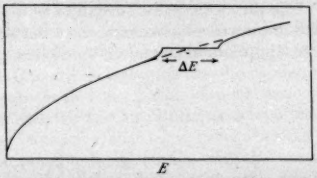


Fig. 2. Relation between energy and wave number for one-dimensional metal with periodic field (broken curve represents curve for free electrons).

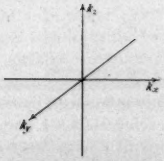


Fig. 3. Axes in k-space.

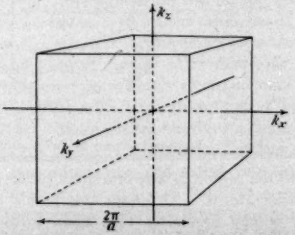


Fig. 4. First Brillouin zone for simple cubic structure.

the wave number k_x at which reflection occurs is equal to $\pi/a \cos\theta$. But, assuming the direction of propagation to lie in the xy plane, $k_x = k_x/\cos\theta$. Our condition thus becomes $k_x = \pi/a$, which is identical with that above (equation (14)) for $n_1 = 1$, $n_2 = n_3 = 0$. The critical values of k, for which discontinuities in the energy occur, are therefore those for which waves with various directions of propagation are reflected by crystal planes, and will clearly vary according to the direction of propagation. For a more complete description of these critical values, it is usual to plot them in a three-dimensional space formed by taking axes k_x, k_y and k_z mutually at right angles (fig. 3). The critical values of the vector $\mathbf{k}$ are then plotted, and fall on planes in the so-called k-space; the planes enclose volumes within which the energy may vary continuously, but on crossing the planes there is a discontinuity in the energy. Figure 4 shows the first volume for a simple cubic lattice; the planes correspond with reflection of waves of wavelength

$2\pi/k$ from the {100} planes, and lie at a perpendicular distance π/a from the origin. Outside this zone, bounded by planes parallel to the {110} planes of the crystal, is another volume, within which the energy may vary continuously. The planes of energy discontinuity which enclose this second volume lie at a perpendicular distance $\sqrt{2}\pi/a$ from the origin, that is, at the same distance as the edges of the cubic volume shown in fig. 4. The relationship between the two volumes is shown in fig. 5. Further volumes within which energy can vary continuously, bounded by planes of energy discontinuity, lie outside these figures. The hypothetical 'k-space', therefore, may be regarded as divided up into successive *zones*, which are termed Brillouin zones because of their detailed examination by L. Brillouin.

The assumption that the potential V is small compared with the total kinetic energy of the electrons (approximation of nearly free electrons) is not necessarily true for real metals. The theory may also be applied to cases where the atoms are well separated, so that the interaction between them is small. This treatment is usually known as the approximation of 'tight binding', the behaviour of an

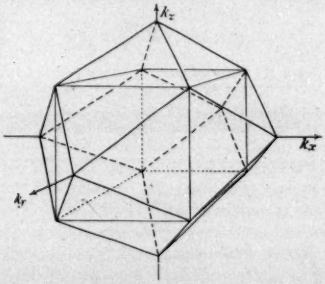


Fig. 5. First and second Brillouin zones for simple cubic structure.

electron in the region of any one atom being only slightly influenced by the field of the other atoms. Although the details of the development are somewhat different, it may be shown that again forbidden bands of energy may occur; the final result is the same as for the former approximation. Considering a simple cubic structure, it is found that the energy of an electron may be written

$$E_0 - \alpha - 2\gamma(\cos k_x a + \cos k_y a + \cos k_z a). \qquad \ldots\ldots(15)$$

where a is the lattice spacing. In this expression, α is an integral depending on the difference between the potentials in which the electron moves in the lattice and in the free atom, while γ has a similar significance. The third term in the above expression varies between sharply defined limits, corresponding to the fact that there exists a band of energies in the crystal. The integral γ, which defines the breadth of the band, is greater the more the wave functions overlap. For the valency electrons of real metals, however, the overlap is too great for the method to give accurate results, though valuable qualitative information may be obtained. For real metals, which represent a case intermediate between the two approximations referred to, a method developed by Wigner and Seitz (1933), which is referred to in detail below, is more appropriate.

The variation of energy as a function of k within the Brillouin zones, and the variation of $N(E)\,dE$ with energy, are of considerable importance for the understanding of real metals. Without considering these factors in detail, however, we may see how the model outlined above accounts for the existence of metallic properties, insulators, and semiconductors. We have seen that, when an electric current is carried, electrons are 'accelerated', that is, promoted to higher energy levels. In order that this may occur, there must be vacant energy levels, above that occupied by the most energetic electron in the absence of an electric field, into which the electrons may be excited. If the value of k corresponding to the highest occupied level is well below the smallest critical value of k at which an energy discontinuity arises, there exist many vacant levels within the first zone into which electrons may be excited. Conduction is therefore possible; this case corresponds with the noble metals copper, silver, gold, and the alkali metals. If, however, all the available energy levels within the first zone are full, and the lowest possible electronic energy at the bottom of the second zone ($E_{\text{min II}}$) is higher than the highest energy in the first zone ($E_{\text{max I}}$) by an amount ΔE, there exist no vacant levels into which electrons may be excited without receiving the prohibitively large increase in energy ΔE. Under these conditions no current can be carried by the material and an insulating crystal results. It may happen that the lowest energy in the second zone is lower than the highest energy in the first zone. It is then possible for electrons, once the energy $E_{\text{min II}}$ is exceeded, to begin to occupy energy levels contained within the second zone, as well as to continue to fill up the vacant levels in the first zone. If the energy of the most energetic electron present exceeds $E_{\text{max I}}$ the first zone will be full and a certain number of levels in the second zone occupied. If, on the other hand, the energy of the highest electron lies between $E_{\text{min II}}$ and $E_{\text{max I}}$ some levels at the bottom of the second zone will be occupied and some at the top of the first zone will be left vacant. In either case, vacant levels exist into which electrons may be excited *without* acquiring the large energy increase referred to in the case of an insulator, and metallic conduction is possible. Magnesium, beryllium, zinc, and other polyvalent metals are materials of this class. For yet another class of crystals, the zone structure is analogous to that of insulators but with a very small value of ΔE. In such cases, at low temperatures the material behaves as an insulator with a high specific resistance; as the temperature rises, however, a small number of electrons are thermally excited across the small energy gap and enter the second zone, where they are free to produce metallic conduction. These substances are termed semiconductors, and their resistance *decreases* with rise in temperature, in marked contrast to the behaviour of real metals. The part played by impurities in such materials is of considerable practical and theoretical interest, and, although it is beyond the scope of the present report, reference may be made to the work of Shockley (1951).

The electronic theory of metals, therefore, has developed very considerably since the time of the Lorentz theory, the major advances being the introduction of quantum principles and the methods of wave mechanics, and the proper attention given to the periodic nature of the potential field in which the electrons move in real crystals. Each advance has resulted in added mathematical complexity, to which little reference can be made in this Introduction, for reasons of space. The major part of this Report, however, will be devoted to a review of the electronic structures of typical classes of metal, and the extent to which certain

physical properties are interpretable on the basis of the present theory. In the course of the review some indication will be given of the bases of the mathematical procedures which have been employed in developing the theory of specific metals.

§ 2. THE ALKALI METALS

The metals of group IA of the Periodic Table are all body-centred cubic in crystal structure and have one valency electron per atom. The Brillouin zone corresponding to this structure is bounded by the planes in k-space which are parallel to the {110} planes in the crystal. Its form is illustrated in fig. 6, and the volume of this zone corresponds with two electrons per atom. There are, therefore, in the alkali metals only sufficient valency electrons per atom to account for half the volume of the zone, and it may be expected that these valency electrons will behave very similarly to free electrons (see fig. 2). To a great extent this is confirmed by the physical properties of these metals, and by the theoretical work which has been carried out.

From the Introduction, it will be appreciated that the approximate physical picture of an alkali metal supposes the valency electrons to be distributed in a

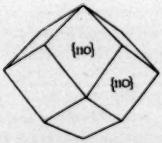


Fig. 6. First Brillouin zone for body-centred cubic structure.

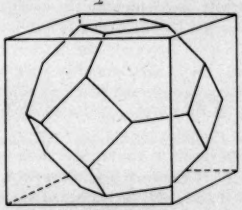


Fig. 7. Atomic polyhedron for body-centred cubic structure.

band of energies; this distribution accounts for the typical properties of these metals. If, therefore, any attempt is to be made to interpret these properties theoretically, it is necessary to calculate the energy corresponding to the bottom of the band, and the average energy of the electrons contained in the band, bearing in mind that these electrons are under the influence of a periodic force-field provided by the metal ions. The problem is complex, and it is of interest to consider the types of approach which have been used. The first necessary step is to establish a sufficiently good approximation to the potential field in which the electrons move; in this Report, we shall not be concerned with this type of calculation, and it will suffice to say that methods analogous to the 'self-consistent field' method used by Hartree (1928) for calculating the wave functions of electrons in free atoms are capable of giving a usable result. It is then necessary to find the wave function, and hence the energy, corresponding to an electron in the lowest energy state of the valency band. The problem was first solved, to a fair degree of accuracy, by Wigner and Seitz (1933, 1934), who considered the metal to be, in effect, divided up into polyhedral cells, each containing one atom. Each cell is obtained by drawing planes perpendicular to, and bisecting, the lines joining each atom to its nearest and next nearest neighbours. Repetition of the resulting polyhedron, which is illustrated in fig. 7, can effectively fill up the whole of space. At the centre of each cell is situated the singly charged positive ion; the potential

energy of the electron within the cell is assumed to be spherically symmetrical and is denoted by $V(r)$, r being the distance from the cell centre. $V(r)$ is taken as the potential of the free singly charged ion. The wave function for the electron in its lowest state is periodic with the period of the lattice; Wigner and Seitz assume, therefore, that the wave function for the electron in its lowest state joins up smoothly at the boundary of each atomic cell. This condition may be summarized by the expression

$$\partial\psi/\partial n = 0 \qquad \ldots\ldots(16)$$

where the differentiation is normal to any bounding plane of the cell. The use of this condition introduces certain complexities into the calculations, and, as a close approximation, it is considered that the atomic cell may be replaced by a sphere of equal volume, when the condition becomes

$$(\partial\psi/\partial r)_{r=r_s} = 0 \qquad \ldots\ldots(17)$$

where r_s is the radius of the sphere. The wave function within each sphere is now obtained by solving the Schroedinger equation:

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{d\psi}{dr}\right) + \frac{2m}{\hbar^2}[E_0 - V(r)]\psi = 0, \qquad \ldots\ldots(18)$$

subject to the appropriate boundary condition, where E_0 is the energy of an electron in its lowest state. In essence, the calculation consists of determining, for an assumed value of E_0, the value r_s of r for which $\partial\psi/\partial r$ vanishes. The determined

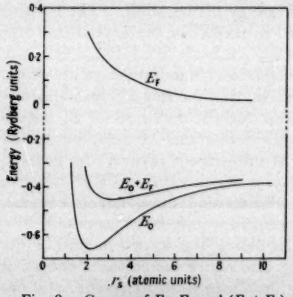


Fig. 8. Curves of E_0, E_T and (E_0+E_T) against r_s for sodium.
(*After* Wigner and Seitz 1934)

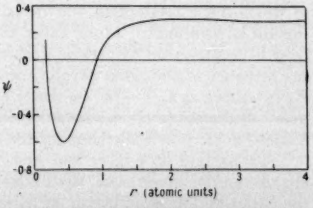


Fig. 9. Wave function of electron in lowest state for sodium.
(*After* Wigner and Seitz 1934)

wave functions then take a form which, close to the ionic nucleus, is very similar to that of the wave function of the free atom. At distances from the nucleus greater than that corresponding to the effective ionic radius, however, there are in general, for a given value of E_0, two values r_s of r which satisfy the boundary condition. As the energy E_0 is decreased, these two values approach and finally disappear. Thus the energy E_0 of the electron in its lowest state may be determined as a function of r_s, as illustrated for sodium in fig. 8. Figure 9 shows the appropriate wave function, calculated at the minimum value of E_0, for sodium, and it will be noted that this function is practically constant for the greater proportion of the atomic volume. This confirms that, in the alkali metals, the valency electron behaves very much like a free electron in a uniform potential field.

The problem now remains of determining the mean energy of the electrons within the energy band. If the wave function for the electron in its lowest state is denoted by $\psi_0(r)$, then it may be shown that the wave function for higher states may be written, within any one atomic sphere, $\psi_k(r) = e^{ikr}\psi_0(r)$. This approximation is only appropriate when, as for the alkali metals, $\mathbf{k}$ is not near the boundaries of the first Brillouin zone, and leads to

$$E_k = E_0 + \hbar^2 k^2/2m. \qquad \ldots\ldots(19)$$

It is thus equivalent to neglecting the periodic field. The mean energy of the conduction electrons, measured from that of the bottom of the band, is therefore, as noted above, $\frac{3}{5}E_{max}$; since λ, the wavelength associated with the electron, is equal to h/mv, and $k = 2\pi/\lambda$, the mean energy of the conduction electrons $E_F = 3\hbar^2 k_{max}^2/10\,m$. The maximum value of the wave number, if there are N electrons present per unit volume, may be shown to be $k_{max} = (3\pi^2 N)^{1/3}$. Putting $N = (4\pi r_s^3/3)^{-1}$, we obtain an expression for E_F in terms of r_s

$$E_F = \frac{3}{10}\frac{\hbar^2}{m r_s^2}\left(\frac{9\pi}{4}\right)^{2/3}. \qquad \ldots\ldots(20)$$

This function is also shown plotted in fig. 8. The sum of the curves for E_0 and E_F, which is shown as the third curve, denotes to a fair approximation the total electronic energy, so that the binding energy per atom, relative to the free atom, is $-(E_0 + E_F + E_I)$ where E_I is the ionization energy of the atom. The binding energy is usually defined as the work needed to dissociate the crystal into free neutral atoms.

The total energy so obtained depends on many factors, one of the most important of which is the assumption, for the excited states, that the free electron approximation is legitimate. More exact expressions for the variation of E_F with k have been suggested by various workers, but the derivations are complex and the mathematics is somewhat beyond the scope of the present review. In general, E_k is written as $E_0 + \alpha\hbar^2 k^2/2m$ where α is a factor which expresses the 'effective mass' of the electron. $\alpha = (m/\hbar^2)(d^2E/dk^2)_{k=0} = \text{electronic mass/'effective mass'}$. Various approximate methods for the assessment of α have been proposed.

The method of Wigner and Seitz for calculating the wave function of the electron in the lowest state, the Fermi energy, and hence the cohesive energy of the lattice and its equilibrium lattice spacing, has been most successful, and has provided a general method from which further developments have been made. It is important to detail some of the assumptions and approximations which are involved, particularly with regard to the cohesive energy. The potential energy of the lattice is regarded as the interaction between the charges in one atomic cell, together with the interaction of the polyhedra with each other. It is now assumed that each polyhedron contains only one electron, so each electron moves in the field of the singly charged ion situated at the centre of the cell. Each cell is thus electrically neutral, so that the interaction between cells is in any case small; on this assumption, the calculation reduces to finding the energy of the charges within an isolated cell, which is further assumed to be spherical. The total energy of the crystal is then the sum of the kinetic energy of each electron, and of its potential energy in the field of its own positive ion. It may be noted that this treatment neglects correlation between the positions of the electrons. Various

estimates have been made of the effect of this neglect, and it may be shown that the necessary net corrections are small. Allowance must be made for the fact that each electron moves not only in the field of the ion, but also in that of a negative charge distribution due to other electrons. Further, electrons with parallel spin quantum numbers tend to keep apart, as do, to some extent, electrons with antiparallel spins. It has been estimated that the correcting terms are as follows: (*a*) interaction energy between electrons $0{\cdot}6\,e^2/r_s$, (*b*) correction for parallel-spin electrons keeping apart $-0{\cdot}458\,e^2/r_s$, (*c*) correction for antiparallel-spin electrons keeping apart $-be^2/r_s$, where b is a quantity which depends upon the value of r_s. The total, $(0{\cdot}14-b)e^2/r_s$, gives the increase in total energy, and is small (of the order of 0·2 ev. for sodium), so that the effect of neglecting these terms is not serious (see below). The Wigner–Seitz calculation, as noted above, replaces the polyhedral atomic cell by a sphere; the effect of this, for sodium and the other alkali metals, is negligible, and is of the order 0·03 ev.

The general method of attack employed by Wigner and Seitz has been extended by Slater (1934) in an effort to obtain more accurate estimates of the energies of the excited electrons, and a brief description of Slater's method may serve to indicate the complexity of the problem. Wigner and Seitz consider in detail only the lowest level, and use only one wave function derived from the s level of the free atom; the improvement introduced by Slater is to consider not one, but eight separate wave functions in combination. The boundary conditions for an arbitrary electron momentum are then to be satisfied at the mid-points of lines joining an atom to its eight nearest neighbours. Solution is carried out for an arbitrary direction of propagation in one of the principal planes, and wave functions and energy levels are determined as functions of internuclear distance as before. Since eight wave functions are needed, Slater used one s function, three p functions, three d functions and one f function. It may be noted that the method does not reproduce correctly energy levels originating in any atomic level except the s and p levels, because two of the five d functions, and six of the seven f functions, are omitted.

Slater considers a wave which is modulated according to the plane wave function $e^{i\mathbf{kr}}$. Then, according to Bloch's theorem, on passing from one point in the lattice to the equivalent point in another cell, the wave function is multiplied by the factor $\exp(i\mathbf{k}\mathbf{r}_{ij})$, where $\mathbf{r}_{ij}$ is the vector from nucleus i in one cell to nucleus j in the other, that is, the perpendicular distance between opposite faces of the cell. The boundary condition employed is then that, on passing from a point on the surface of one cell to a perpendicularly opposite point, the wave function must be multiplied by $e^{i\mathbf{kr}}$, where $\mathbf{r}$ is equal to the perpendicular distance between faces. Each of the three components of the gradient is, of course, multiplied by the same factor. The spherically symmetrical wave equation is now solved, subject to this condition on the surface of a polyhedral cell. With the eight wave functions used, the boundary conditions are applied at the centres of the hexagonal faces of the atomic cell. If u is the wave function considered, Slater regards it as composed of an even part u_g, which is unchanged on passing from one point to that diametrically opposed, and an odd part u_u, whose sign changes. Then, if the function is u_g-u_u at one point, it is u_g+u_u at the other. Similarly, the normal derivatives are $u_g'-u_u'$ and $u_g'+u_u'$, and the boundary conditions are:

$$u_g+u_u=e^{i\mathbf{kr}}(u_g-u_u) \quad \text{and} \quad u_g'+u_u'=-e^{i\mathbf{kr}}(u_g'-u_u') \quad \ldots\ldots(21)$$

where the u values are calculated at the mid-points of faces, $\mathbf{r}$ is the vector joining mid-points, and there are, for the body-centred cube, four such conditions for the four pairs of opposite faces. Thus

$$\frac{u_g}{u_u} = -\frac{1+e^{i\mathbf{kr}}}{1-e^{i\mathbf{kr}}} = -i\cot(\tfrac{1}{2}\mathbf{kr}), \qquad \ldots\ldots(22)$$

while

$$u_g'/u_u' = i\tan(\tfrac{1}{2}\mathbf{kr}). \qquad \ldots\ldots(23)$$

If $|u_u|$ is the absolute magnitude of u_u, $u_u = i|u_u|$, and the equations may be combined in the form

$$|u_u|/u_g = -u_g'/|u_u|' = \tan(\tfrac{1}{2}\mathbf{kr}). \qquad \ldots\ldots(24)$$

The four equations of this type give eight equations in all to be satisfied for the boundary conditions, corresponding respectively to the one s, three p, three d, and one f wave functions referred to above. For details of the somewhat complicated handling of the necessary mathematics, Slater's original paper should be consulted; it should be noted that the eight equations which have to be satisfied are too difficult to handle in the general case, and it is necessary to limit the calculations to the case of propagation in an arbitrary direction in the xy plane. These results are

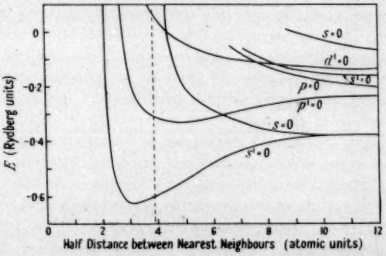


Fig. 10. Energy band system for sodium, as a function of distance between atoms.
(*After* Slater 1934)

sufficiently general to indicate the essentials of the general solution, and are of considerable interest. Using the same potential curve for sodium as that used by Wigner and Seitz, the s, p, d and f wave functions corresponding to a variety of different energies are determined; from these a diagram may be constructed, showing electronic energies as functions of internuclear distance, as illustrated in fig. 10. In this figure, the curves shown are those for which $s'=0$, $p'=0$, etc., and $s=0$, $p=0$, etc., where the symbols s and p denote the radial parts of the corresponding functions and s' and p' the normal derivatives. These curves divide up the area of the figure into different regions, within which the wave equations have a definite number of real solutions. Thus the region between the curves $s'=0$ (which corresponds with the 'ground state' curve of Wigner and Seitz, fig. 8) and $s=0$ derived from the 3s state of the free atom of sodium corresponds with a band of energy states which may be occupied by electrons. Similarly for the area defined by the curves $p'=0$ and $p=0$ derived from the 3p level of the atom. For large distances the allowed regions are narrow bands surrounding each energy level of the free atom, but the bands broaden as the atoms come together until they touch or overlap. It will be seen that at the equilibrium

distance apart of atoms, which is denoted by a dotted line, overlap between the 3s and 3p bands is considerable; it is only at high energies, however, that any contribution occurs from the 3d band. This diagram illustrates the fact that the conduction electrons in the valency band are not necessarily all derived from the same atomic level. It may also be noted that, in Slater's treatment, only three states originate from each d level, instead of five, because only eight functions were used to set up the approximate solution. There are, therefore, missing states in the calculated electronic distribution, but the error introduced by this circumstance is only likely to be important at the higher energies. A more reliable solution would require the use of, say, fourteen functions, which would include the eight already used and also six satisfying boundary conditions at the mid-points of the square faces in fig. 7.

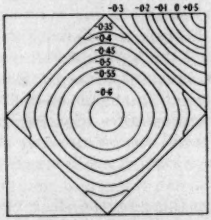


Fig. 11. Energy contours in k-space for sodium ; k_x k_y plane.
(*After* Slater 1934)

Fig. 13. Wave function corresponding to lowest energy in the first Brillouin Zone for sodium, plotted for propagation in the [110] direction.
(*After* Slater 1934)

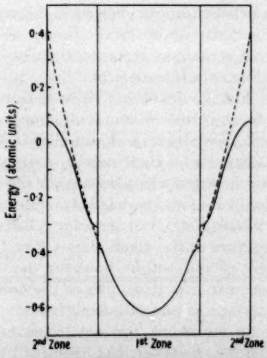


Fig. 12. Relation between energy and wave number for the [110] direction in the crystal of sodium.
(*After* Slater 1934)

Although, in fig. 10, forbidden energy ranges appear between the permitted bands at large distances of separation, they are not evident at distances approximating to the interatomic distance. This is because the calculations upon which fig. 10 is based consider, effectively, all possible directions for $\mathbf{k}$, the wave vector of the electrons. The energy gaps which separate the first and second Brillouin zones become clear if a particular direction of propagation is considered. This is illustrated by fig. 11, which is due to Slater, and shows contours of constant energy in the $k_x k_y$ plane of k-space. Along the 45° diagonal, it will be noted that there is no contour for energy $-0{\cdot}35$ atomic units; instead, the energy jumps from approximately $-0{\cdot}37$ units to $-0{\cdot}32$ units. This may also be seen in the plot of energy against wave number (fig. 12), for the [110] direction in the crystal, which

corresponds to the 45° diagonal in fig. 11. The corresponding curve for free electrons is shown as a broken curve, which is a parabola; the agreement between the two curves within the first Brillouin zone is extremely close. For the lower portion of the zone, which alone is filled with electrons in the monovalent alkali metals, there is little difference between the two curves and the electrons thus behave as if effectively free. This correspondence with free electron characteristics is also reflected in the close approximation of the lower energy contours in fig. 11 to circles, showing that the energy depends only on the magnitude of the momentum and hardly upon its direction, except in the vicinity of zone boundaries. The forms of the wave functions themselves are also of interest. It will be noted that the lowest energy in the first zone is approximately $-0{\cdot}60$ atomic units at the normal distance of separation. Figure 13 shows the wave function corresponding to this energy, and representing propagation in the [110] direction. The line forming the horizontal axis is drawn through a close-packed line of atoms in the [111] direction, which is the closest direction to the wave normal in which such a close-packed row can be found. The real part of the wave function is plotted as a function of distance along the line joining nuclei, and it may be noted that, near nuclei, the wave function has the character of an atomic wave function. In between nuclei, however, the wave function approximates very closely to the broken sinusoidal curve; thus in all parts of the metal where the valency electron is not actually penetrating the core of an atom, its wave function is closely similar to that of a free electron.

The results obtained for sodium are typical of those for the alkali metals in general. Other methods of computation, differing in detail from those described above, have been used, but all investigations confirm that the free-electron picture is adequate for these metals, as regards both electronic energy and the part of the wave function which lies outside the atom. The computational methods developed allow a reasonably detailed investigation of the nature of the energy gaps and the Brillouin zones, but accuracy is lost in the higher energy regions. The electronic structure of the alkali metals may thus be regarded as satisfactorily established; there are indications, however, that higher accuracy is needed, in so far as calculation from first principles of the cohesive energies and lattice spacings are not in exact agreement with experiment.

Of the alkalis, the metal sodium has been examined most fully; considerable attention, however, has been paid to lithium, for which it appears that the free electron approximation is somewhat less satisfactory. Whereas for sodium the effective mass of the electron is near to unity, in the case of lithium it is greater. For this metal the value of α discussed above in considering the Fermi energy contribution to the electronic energy is of the order of 0·75 (Millman 1935).

An indication of the degree of agreement between calculated and experimentally observed values of the binding energies and lattice spacings of the alkali metals may be obtained from the following comparisons (Seitz 1940). In deriving

Metal	Lattice spacing (Å)		Binding energy (kcal/g atom)	
	Calculated	Observed	Calculated	Observed
Li	3·50	3·46	36·2	39
Na	4·51	4·25	24·5	26
K	5·82	5·20	16·5	23

these figures, the minimum in the final curve of energy against r_s, after making the corrections referred to above, was taken to define the theoretical lattice spacing,

and the cohesive energy was calculated for this particular value of r_s. If the cohesive energies are calculated for the observed lattice spacings, the following values are obtained: Li 36·1 kcal/g atom, Na 23·5 kcal/g atom, K –.

It will be seen that, for lithium and sodium, agreement is reasonable; for potassium, however, the calculated lattice spacing is much too high, and the cohesive energy too low. The curve of energy against r_s is such as to give a negative cohesive energy at the observed lattice spacing. The results for potassium are thought to be in error on account of imperfections in the potential field upon which the calculations are based (Gorin 1936).

It is instructive to note the large contributions made to the cohesive energy by the exchange and correlation energy terms. Without correction, the cohesive energy of the crystal relative to that for free atoms may be written $(E_0+E_F+E_I)$ where E_I is the ionization energy. For lithium, this is $-39{\cdot}0$ kcal/g atom, which is in itself numerically close to the observed binding energy.

The correction terms are as follows: (*a*) interaction energy between electrons 114·8 kcal/g atom, (*b*) exchange energy correction $-90{\cdot}2$ kcal/g atom, (*c*) correlation energy correction $-21{\cdot}7$ kcal/g atom. The total energy is thus $-36{\cdot}1$ kcal/gn atom and corresponds with the observed binding energy value of 39 kcal/g atom. The two negative terms almost balance the positive term, and errors in the estimation of the individual corrections could therefore introduce a significant error into the final values. The error for lithium of approximately 3 kcal/g atom is probably due to errors in the calculation of the correlation energy.

It is of some interest to consider how the band structure of the alkali metals interprets their magnetic properties. This question has been investigated by Sampson and Seitz (1940) whose work serves to indicate the extent to which it is possible to derive quantitative information about physical properties from the theory in its present state. According to the original Pauli theory of paramagnetism, in which exchange and correlation interactions between the electrons are neglected, the volume susceptibility is

$$\chi_p = 2\beta^2 g_s / V \qquad \ldots\ldots(25)$$

where β is the Bohr magneton, g_s is the density of electronic levels of one spin at the top of the filled band, and V is volume of the specimen. For the diamagnetic susceptibility, considering free electrons,

$$\chi_d = -4\pi m\beta^2 h^{-2}(n/3\pi V)^{1/3} \qquad \ldots\ldots(26)$$

according to the work of Landau (1930), which has been extended by Peierls (1933) and Wilson (1936) to include effects consequent upon the electrons not being perfectly free. n is the total number of electrons.

In the case of the paramagnetic susceptibility, some allowance may be made for the fact that the electrons are not entirely free by assigning to the electron an effective mass m^* in place of the electron mass m. The paramagnetic susceptibility may then be written

$$\chi_p = 3Nm^*\beta^2/\hbar^2 k_0^2 \qquad \ldots\ldots(27)$$

where k_0 is the value of k at the top of the filled levels, and N is the number of electrons per unit volume. Using values of m^*/m derived by Bardeen (1936) for sodium and lithium, the values of χ_p for these two metals are respectively 0·59 and 1·22 c.g.s. units $\times 10^{-6}$. The inclusion of exchange and correlation terms by Sampson and Seitz has the effect of taking into account the change in

the effective atomic mass as electrons are redistributed among the energy levels under the influence of a magnetic field.

If the metal has N electrons per unit volume, then in a given state of magnetization there are $(\frac{1}{2}N-p)$ electrons associated with one spin and $(\frac{1}{2}N+p)$ associated with the opposite spin. When p is zero, as noted above, the energy of interaction of electrons of parallel spin is $-0{\cdot}458\,e^2/r$ where $4\pi r^3/3 = 1/N$. When p is not zero, it may be shown that the average values of this interaction energy for the two half-bands are respectively $-0{\cdot}458\,xe^2/r$ and $-0{\cdot}458\,ye^2/r$, where

$$x=[(N+2p)/N]^{1/3} \quad \text{and} \quad y=[(N-2p)/N]^{1/3}.$$

The total exchange energy is $-0{\cdot}458\,e^2(x^4+y^4)N/2r$. The correlation energy of interaction between electrons of opposite spin, according to Sampson and Seitz, is:

$$-e^2\{[(\tfrac{1}{2}N+p)f(r/y)]+[(\tfrac{1}{2}N-p)f(r/x)]\}. \qquad \ldots\ldots(28)$$

The paramagnetic susceptibility is then expressed as

$$\chi_p = N\beta^2/2\alpha \qquad \ldots\ldots(29)$$

where α is a quantity which includes the exchange and correlation corrections. From this equation, the paramagnetic susceptibility is calculated as 3·54 and 1·11 c.g.s. units $\times 10^{-6}$ for lithium and sodium respectively. The corresponding α values are 0·35 and 0·61. The following are the contributions made by the exchange and correlation terms to the value of α:

Metal	Uncorrected	Exchange correction	Correlation correction
Li	1·02	−0·86	0·19
Na	1·12	−0·70	0·19

It may be seen, therefore, that the refinements to the theory result in a reduction of α for lithium by a factor of nearly 3; the correction for sodium is somewhat less serious.

The corrections to be applied to calculation of diamagnetic susceptibilities were also discussed by Sampson and Seitz, but, in view of the complexity of the problem, no actual corrections were applied to the numerical values of $-0{\cdot}17$ and $-0{\cdot}23$ c.g.s. units $\times 10^{-6}$ for lithium and sodium respectively which may be derived from the extension of the free electron theory by Peierls. Diamagnetism resulting from the interaction of the cores of the ions was also considered, and shown to be small. The numerical values calculated (in c.g.s. units $\times 10^{-6}$) are summarized in the following table:

Metal	Spin paramagnetism	Free electron diamagnetism	Ion core diamagnetism	Total	Experimental
Lithium	3·54	−0·17	−0·05	3·32	2·0 (*a*) 1·4 (*b*)
Sodium	1·11	−0·23	−0·18	0·70	0·59 (*c*) 0·59 (*d*) 0·63 (*e*)

(*a*) Bitter (1930); (*b*) Rao and Sririman (1937); (*c*) Sucksmith (1926); (*d*) McLennan, Ruedy and Cohen (1927); (*e*) Lane (1929).

The agreement for sodium is fair, and supports the general conclusion that the valency electrons behave as if free. On the other hand, the theoretical value for

lithium is too large, and suggests that the calculations give too much emphasis to the paramagnetism. This work shows that, although the general theory of the alkali metals is in a satisfactory state, there is still some difficulty in the quantitative derivation of physical properties, particularly where the property is sensitive, as in the case of the magnetic susceptibility, to the exact value of the effective electronic mass.

The theoretical investigations carried out on the alkali metals allow estimates to be made of the elastic properties of these materials. Adopting the notation of Voigt (1928), we may write the strain energy of a crystal as follows:

$$W = \tfrac{1}{2}\Sigma\Sigma C_{ij}e_ie_j \qquad \ldots\ldots(30)$$

where $e_1, e_2, \ldots . e_6$ are the six components of the strain, and C_{ij} represents the elastic constants. For cubic crystals, the elastic properties may be described in terms of the three constants C_{11}, C_{12} and C_{44}; for an isotropic medium, C_{44} is equal to $(C_{11}-C_{12})/2$. When a crystal is elastically strained, the energy changes, and it may be shown that

$$\frac{\partial W}{\partial e} = \frac{1}{2}\frac{\partial^2 W}{\partial e^2}\,de, \qquad \ldots\ldots(31)$$

so that the quantity $\frac{1}{2}\partial^2 W/\partial e^2$ forms the appropriate elastic constant. Calculation of the strain energy for certain specified elastic deformations will thus allow estimates to be made of the elastic constants.

The energy of the body-centred cubic crystal has been considered above. The effect of elastic deformation, except for the particular deformation corresponding to a uniform hydrostatic compression or expansion, is to distort the atomic cells into which the metal is considered to be divided. It is thus necessary to consider the influence of distortion on the various terms into which the energy may be subdivided. According to Fuchs (1936) and later workers, the kinetic energy of the lowest state of the electron may be taken as a function of r_s alone where r_s is the radius of the atomic cell. It is not appreciably changed, therefore, by a distortion which does not alter the volume. This assumption is justified by the fact that the wave function for the lowest electronic state ψ_0 is practically constant near the boundary of an atomic polyhedron. Further, since the Fermi energy of free electrons depends only upon volume, it is justifiable to assume a similar dependence for the electrons in alkali metals. There remain the electrostatic 'coulomb' interaction energy of the electrons and ions, the exchange and correlation energies, and the non-coulomb ion–ion interaction. As for free electrons, the exchange and correlation energies may be assumed to depend only upon volume, while for alkali metals, the contribution to the cohesive energy by the ion–ion interaction is very small. The main term in the determination of the elastic constants will thus be the coulomb interaction between ions and electrons. This may be calculated; the charge density near the edges of the atomic polyhedron is approximately $-e/\Omega_0$, where Ω_0 is the volume per atom, so that the change in the potential energy of the valency electrons on distortion is the same as that of a lattice of positive point charges $+e$ embedded in a uniform negative charge distribution $-e/\Omega_0$.

Since no distortion of the atomic polyhedra is involved, the compressibility is the simplest of the elastic properties to calculate; it is determined by the curvature of the curve of E against r_s. The pressure applied may be written

$p=-\partial E/\partial v$ where E is the internal energy per atom. Thus $1/\beta$, where β is the compressibility, may be related to the energy by the following expression:

$$\frac{1}{\beta}=-v\frac{\partial p}{\partial v}=v\frac{\partial^2 E}{\partial v^2}. \qquad \ldots\ldots(32)$$

From the theoretical calculations, the variation of E with r_s is known. Since $v=4\pi r_s^3/3$:

$$\frac{1}{\beta}=\frac{r_s}{3}\frac{\partial}{\partial r_s}\left\{\frac{1}{4\pi r_s^2}\frac{\partial E}{\partial r_s}\right\}=\frac{1}{12\pi r_s}\frac{\partial^2 E}{\partial r_s^2}. \qquad \ldots\ldots(33)$$

The compressibility of lithium and sodium has been computed from this relationship for a range of values of r_s, and good agreement with experiment is obtained for sodium. The agreement for lithium is less good, but, as seen earlier, the calculations for this metal are in general less consistent with experiment than those for the other alkalis. The compressibility β is related to the elastic constants C_{11} and C_{12} by the expression

$$1/\beta=\tfrac{1}{3}(C_{11}+2C_{12}). \qquad \ldots\ldots(34)$$

For the further study of the elastic properties of the alkali metals, two specific types of dilation have been considered. For the case of equal contraction and expansion along two cube axes, leaving the volume unchanged, the elastic constants are related to the energy change by the expression

$$C_{11}-C_{12}=\frac{1}{2\Omega_0}\frac{\partial^2 E}{\partial \epsilon^2}, \qquad \ldots\ldots(35)$$

where ϵ is the strain ($+\epsilon$ in the x direction, $-\epsilon$ in the y direction). $C_{11}-C_{12}$ is thus calculable in terms of the change in the energy of the interaction of the electrons and the ions. For a shearing strain in a plane parallel to two cube edges, the compressional strains are zero, and

$$C_{44}=\frac{1}{\Omega_0}\frac{\partial^2 E}{\partial \gamma^2}, \qquad \ldots\ldots(36)$$

where γ is the angle of shear. Again, therefore, C_{44} is calculable. It may be noted that these distortions, unlike that which determines the compressibility, distort the atomic spheres into ellipsoids.

The results of such calculations are of considerable interest. In the case of sodium, which has been the most fully studied, the contributions to $C_{11}-C_{12}$ made by the coulomb term and the ion–ion interaction term are respectively $0{\cdot}143\times10^{11}$ and $-0{\cdot}02\times10^{11}$ dyn/cm^2. In agreement with the fact that the ion overlap is very small for the alkali metals, this second term is also very small. The total, $0{\cdot}141\times10^{11}$ dyn/cm^2, compares favourably with the experimental quantity of $0{\cdot}145\times10^{11}$ dyn/cm^2. For C_{44} the respective contributions are $0{\cdot}532\times10^{11}$ and $0{\cdot}048\times10^{11}$ dyn/cm^2, making a total of $0{\cdot}580\times10^{11}$ dyn/cm^2 (experimentally observed $0{\cdot}59\times10^{11}$ dyn/cm^2). Again, the contribution due to ion–ion interaction is small. One of the most important aspects of the work is the indication that real metals are far from isotropic; the ratio of C_{44} to $(C_{11}-C_{12})/2$, which should be unity for an isotropic medium, is approximately 8·1 for sodium.

Results for the other alkali metals are similar, and confirm that the coulomb contribution is much larger than the contribution from ion–ion interaction. The agreement with experiment, in cases where this may be tested, gives considerable confidence in the general reliability of the model employed for the calculations.

§ 3. METALS OF GROUP IB OF THE PERIODIC TABLE

Like the alkali metals, these elements are monovalent and typically metallic. They are, however, face-centred cubic in structure. The Brillouin zone is bounded by planes parallel to the {200} and {111} planes of the crystal (fig. 14), and can contain a maximum of two electrons per atom. Again, therefore, there are insufficient electrons per atom to do more than half fill the zone, and the electrons in the metal may be expected to be relatively free. According to this picture, it would be expected that calculations of the Wigner–Seitz–Slater type would give an adequate interpretation of the cohesion, lattice spacing and compressibility of these metals. There is, however, a complication, in that the ions of these metals are relatively much larger than those of the alkali metals, and occupy a larger proportion of the atomic volume. As a consequence, allowance must be made for the repulsion which arises between closed shells of electrons.

The atomic polyhedron for a face-centred cubic metal is shown in fig. 15. Applying to this polyhedron very similar considerations to those described above, it may be shown that the wave function for the valency electron in its lowest

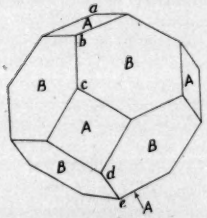


Fig. 14. First Brillouin zone for face-centred cubic structure.

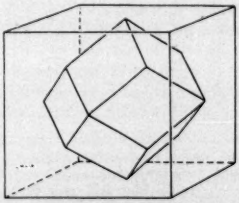


Fig. 15. Atomic polyhedron for face-centred cubic structure.

state takes the form of fig. 16. The oscillations near the nucleus (there are three zeros, since the electrons involved are 4s electrons) may be seen to extend relatively far into the space between nuclei, but again it may be observed that the wave function is relatively flat in the inter-ionic space. This wave function was derived by Fuchs (1935), who, in assessing the potential in which the electron moves, took into account the energy of interaction between the valency electron and the inner shells of electrons. From calculations of the Wigner–Seitz–Slater type, the energy of the lowest electronic state may be derived as a function of r_s, and is shown for copper in fig. 17. Calculation of the Fermi energy by methods somewhat analogous to those used by Wigner and Seitz show its variation with r_s. The total energy of the valency electrons is thus represented by the sum of these curves. The minimum occurs at an r_s value corresponding to a calculated lattice spacing which is in reasonable agreement with that observed. It is, however, as noted above, necessary to take into account the repulsion between the ions. This energy of repulsion has been calculated by Fuchs, and is also included, as a function of r_s, in fig. 17. The result of this is to give a curve, showing the total energy of the metal, for which the minimum occurs at a value of r_s somewhat greater than that corresponding to the observed lattice spacing, but still in fair agreement with experiment. It may be noted that the calculated lattice spacing

(i.e. the r_s value corresponding to the minimum in the curve representing the total energy of the metal) is controlled almost entirely by the ion–ion repulsion term. For copper, therefore, and presumably also for silver and gold, the structures may be regarded as composed of ions in contact; in this case the ions appear to behave very much like hard incompressible spheres. It is this factor which accounts for the close-packed nature of the structures assumed by copper, silver and gold.

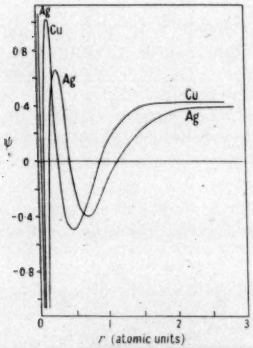


Fig. 16. Wave functions of electron in lowest state for copper and silver. (*After* Mott and Jones 1936)

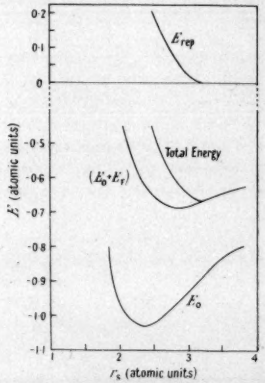


Fig. 17. Curves of E_0, (E_0+E_F), and total energy against r_s for copper. The curve E_{rep} denotes the energy of repulsion between ions. (*After* Fuchs 1935)

For copper, which is a much more complicated case than that of the alkali metals, the observed and calculated values of the binding energy and lattice spacing are as follows :

	Calculated	Observed
Binding energy (kcal/g atom)	33	81
Lattice spacing (Å)	4·2	3·6

The agreement, particularly for the cohesive energy, is thus considerably less good than for the alkali metals.

For the metals of group IB of the Periodic Table, therefore, the physical picture is one of hard spherical ions held together by the electrostatic attraction between the ions and the valency electrons. The valency electron energies lie completely within the first Brillouin zone, which is approximately half-filled. The metals are thus good conductors of electricity, since there is no lack of vacant electronic levels to which electrons may be accelerated.

It is of interest to compare this theoretical picture with that which may be obtained from experimental observations. Of the methods available, the measurement of the absorption of these metals in the ultra-violet region is perhaps the most revealing, in so far as it gives information with regard to the energy levels of electrons of intermediate binding energy. The optical constants of metals in

the visible and near ultra-violet are, of course, also important in providing information about the energy levels of outer electrons. The absorption of gold and silver in the ultra-violet region has been investigated by Simons (1943), who used a specially constructed vacuum spectrograph. In principle, if I_0 is the incident intensity and I the transmitted intensity, then $I = I_0 e^{-Kd}$ where K is the absorption coefficient and d is the film thickness. Corrections must be applied for experimental conditions, such as reflection losses, but it is possible to derive K, and thus also k, the extinction coefficient, by the relation $K = 4\pi\nu k$. The actual shape of the absorption bands can only be deduced from the curve, against the frequency ν, of the quantity nk, where n is the refractive index. The refractive index, however, is not known in the far ultra-violet, and no reliable estimate can be made. It may be shown, however, that the characteristic frequencies for absorption may be obtained, to a close approximation, from the maxima on the plot of the extinction coefficient against the frequency; the refractive index $n = 1$ far outside the region of the classical spectral line, increases slowly as the characteristic frequency is approached from the low frequency side, and decreases suddenly at the point of resonance itself, so that the curve of n against ν is somewhat flat, except near the resonance point, and k determines the shape of the curve of nk against ν. For gold, two characteristic frequencies are observed at wave numbers of 32 000 cm^{-1} and 75 000 cm^{-1}. In the case of silver, only one characteristic frequency is observed at a wave-number of 40 000 cm^{-1}.

The outer electronic structures of the free atoms of gold and silver are similar; gold has ten 5d and one 6s electron, while silver has ten 4d and one 5s electron. Gold, however, may be both univalent and trivalent in its chemical reactions, while silver is uniformly univalent. Thus the d electrons in gold would appear to be less strongly bound than those in silver. The ultra-violet experiments indicate that, in the metallic state also, the d electrons of gold are less strongly bound than those of silver. In metallic gold, the 5d band is full, and the 6s band (the valency band) is partially full, while the 6p band (excited levels) is empty. The absorption bands at 32 000 cm^{-1} and 75 000 cm^{-1} are identified with the 6s – 6p and 5d – 6p transitions. Neglecting temporarily the breadths of the energy levels, these transitions correspond with energy differences of approximately 4 and 9·25 ev. In the case of silver, the corresponding values are approximately 5 and 16 ev, indicating a greater binding energy of the d electrons in silver. The energy bands for gold may be considered in greater detail, with the aid of further data from absorption experiments. From the work of Goos (1936) it may be inferred that the 5d – 6s transition in gold occurs at 26 000 cm^{-1}, while the breadths of the d and s bands are respectively 3·2 ev (Richtmyer, Barnes and Ramberg 1934) and 5·8 ev (de Sauvenier 1939). The characteristic absorption at a wave number of 75 000 cm^{-1} (9·25 ev) corresponds to transition from the centre of the d band to the centre of the p band. In the case of the 5d – 6s transition, since the s band is, from measurements on the Hall effect, filled for 0·3 of its breadth only, the resonant frequency at 26 000 cm^{-1} (3·25 ev) involves the difference between the energy of levels lying at 0·65 of the breadths of the two bands; this is because the frequency peak is taken as corresponding to transition between the centre of the unfilled portion of the s band and the corresponding point in the d band, according to the usual selection rules. The bottom of the s band thus lies at 1·56 ev above the bottom of the d band. The s–p transition takes place from the centre of the occupied part of the s band to a point 0·15 of the way from the bottom of the p band, so that

this point lies at 6·43 ev above the bottom of the d band or 4·83 ev above the centre of this band. According to this scheme, therefore, a value of $(1{\cdot}6+9{\cdot}25-6{\cdot}43)$ or 4·42 ev corresponds with 0·35 of the width of the p band; the width of this band is thus approximately 12·6 ev. From ultra-violet absorption experiments, therefore, it is possible to obtain a picture of the distribution of the outer energy bands, and the relevant data for gold, according to the above discussion, are summarized in fig. 18(*a*). Since the energy values for the d−s and s−p transitions are very similar, it is possible that they have been wrongly identified; in this case, reversal of the identification leads to the scheme of fig. 18(*b*). The breadth of the p band remains at 12·6 ev, but the relative positions of the d and s bands are somewhat altered. It is clear, however, that the full 5d and partially filled 6s bands overlap, and that, in agreement with theory, the top of the filled portion of the s band lies above the highest level of the d band. Somewhat similar distributions of the d and s bands would be expected to be applicable to copper and silver.

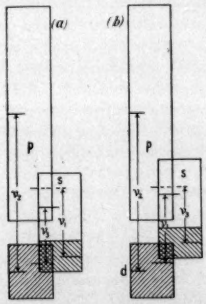


Fig. 18. Diagrammatic representation of the 5d, 6s and 6p bands in gold.
(*After* Simons 1943)

The distribution of the d and s electrons for copper has been studied in detail by Krutter (1935), and this work may be referred to as somewhat complementary to the work of Fuchs. It was carried somewhat further by Slater (1936) and by Rudberg and Slater (1936); the methods used in general followed closely the principles of those already described with reference to the alkali metals. Although the metal copper was considered as the specific case, the same type of distribution may reasonably be expected for silver and gold. As previously, the wave function is taken to be of the form

$$\psi_k(r)=e^{2\pi i(\mathbf{k}.r)}u_k(r) \qquad \ldots\ldots(37)$$

where $\mathbf{k}$ is the wave vector and $u_k(r)$ has the periodicity of the lattice. For the electrons of lower energy, $u_k(r)$ is a rapidly changing function of position and near the lattice points the wave function behaves much like an atomic wave function, as noted above. If known in the neighbourhood of one arbitrary lattice point, the value at any other lattice point may be found; thus, following the Wigner–Seitz technique, Krutter considered a twelve-sided polyhedron with a copper nucleus at the centre. The potential field within such a cell is that of the singly charged copper ion, and the radial part of the Schroedinger equation is solved subject to

the usual boundary conditions, so that the solution in one cell joins smoothly on to the solution in neighbouring cells. Following the treatment of the general problem by Slater, Krutter expresses the wave function in the cell as a linear combination of twelve independent functions, whose angular parts are the twelve lowest spherical harmonics and whose radial parts are the above-mentioned solutions of the Schroedinger equation; coefficients in the linear combination are found by applying the appropriate boundary conditions at the mid-points of the twelve faces. Assuming energy values and the direction of the wave vector, the magnitude of **k** may be calculated; this proves to be sometimes real, and sometimes imaginary, which corresponds to a forbidden energy. In Krutter's calculations, the equations were solved for only three directions in the lattice, but the necessary extension to higher bands was carried out by Slater, and the results were averaged over all directions of propagation. As a result, it was possible to calculate the curve of density of allowed energy states against the energy; this diagram is shown in fig. 19. The shaded part of this diagram indicates the extent to which this band structure is filled by the eleven d and s electrons of copper. The calculations show

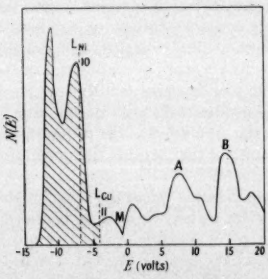


Fig. 19. Density of states plotted against energy for copper.
(*After* Krutter 1935)

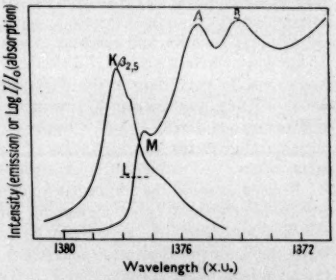


Fig. 20. X-ray absorption and emission curves for copper.
(*After* Beeman and Friedmann 1939)

that the states derived from 3d atomic wave functions are drawn into a band with a high density of states, which is formed from five overlapping 'sub-bands'. The structure extends from -13 to about $-5{\cdot}5$ ev on the energy axis. The 4s band, with a considerably lower density of states, extends according to these calculations from approximately -13 to about $-0{\cdot}5$ ev. At or about the point M, the 4s and 4p bands overlap. It will be noted that there is a general similarity between fig. 19 and the essentials of fig. 18, apart from the energy values involved, and the fact that there appears to be a considerable splitting of the d band in the face-centred cubic structure of copper. The overall accuracy of such a curve is difficult to estimate, but it has been shown that the K absorption limit and $K\beta_{2,5}$ emission lines of copper are not inconsistent with the theoretical results (Beeman and Friedmann 1939).

Using a double crystal precision spectrometer, Beeman and Friedmann determined the absorption limit with considerable accuracy; intensity measurements were made by a Geiger counter technique. The absorption and emission curves obtained are illustrated in fig. 20. The $K\beta_{2,5}$ emission line involves

transitions from the full band to empty K states in the atom, and the high-frequency side of the line closely matches the beginning of absorption in the absorption edge. The decrease in intensity on the high-frequency side of the emission line occurs because beyond a certain point the bands are no longer filled. The absorption band is, of course, due to transitions of electrons into the empty portions of the band system. A knowledge of the density of states in the band is not, in itself, sufficient to give an adequate interpretation of the shape of an emission band, or of the form of an absorption band close to the absorption edge. The intensity of x-ray absorption or emission may be regarded as $I(E) = N(E)\,p(E)$ where $p(E)$ is the electron transition probability, which is also a function of the energy. Although the variation of $p(E)$ is not known in detail, enough may be determined to permit comparison with experiment. Since the transitions involved are between the K level and the outer band structure, the intensity of emission or absorption will depend largely upon the degree to which states derived essentially from the p states of the free atom occur in the band structure. Krutter's calculations show that there is considerable mixing of p function in the 3d and 4s bands, and the position may be summarized by saying that the 3d 'sub-bands' which combine to give the lower energy peak in fig. 19 contain no p function, while the remaining 'sub-bands' do contain p function and combine to give the higher energy maximum. For the 3d band as a whole, therefore, $I(E)$ will have a maximum at about the position of the higher energy maximum in the $N(E)$ curve. It will decrease rapidly to zero at about $-5{\cdot}5$ ev, and less rapidly (owing to a more gradual falling off in the value of $p(E)$) to zero at about -13 ev. In the 4s band, the admixture of the p function is larger, and gives rise to a high value of $I(E)$ in spite of the fact that the density of states is low.

Bearing in mind the somewhat qualitative information about the trend of the transition probability, good agreement with experiment may be obtained between the calculated levels and experiment by identifying the maxima of $N(E)$, A and B in fig. 19, with the absorption maxima A and B in the copper absorption band; the main peak in the emission band corresponds with the maximum of $I(E)$ in the 3d band, while the small lump on the high-frequency side of this maximum corresponds to emission from the filled part of the 4s band which exists above the filled 3d band. With regard to the structure of the absorption band near the edge, the initial portion, up to the small peak M, corresponds with the initial absorption into the 4s band; the decrease in absorption is due to the small overlap between the 4s and 4p bands, while the significance of peaks A and B have already been given. If these correlations are correct, then the following comparison of energy values deduced from the experimental work (fig. 20) and the theoretical work (fig. 19) are possible:

Energy or frequency intervals in figs. 19 and 20	$K\beta_{2,5}-L$	L−M	M−A	A−B
Theoretical energy difference (ev)	3·1	3·4	8·3	7·2
Experimental energy difference (ev)	3·4	4·0	10·7	9·4

The agreement is probably within the uncertainty of the calculations, so that the experimental results give support to the general picture developed of conditions near the high energy limit of the energy bands in copper. The theoretical estimate of the energy width of the d-band would appear, however, by comparison of figs. 18 and 19, somewhat too high.

Since copper, silver and gold each contain one electron per atom, while the volume of the first Brillouin zone contains two electrons per atom, these metals

form particularly convenient solvents with which to study electronic effects on alloying. This subject is referred to later in this review.

It has already been noted, in the section describing the electronic structure of the alkali metals, that the binding energy calculations allow an estimate to be made of the elastic constants of the metal. In the case of copper, since the ions overlap considerably, it may be expected that the contribution from ion–ion interaction will be correspondingly large. This may be denoted by $w(r)$ per ion pair, and is given approximately by the relation

$$w(r) = C \exp\{-(r-2R_i)/\rho\} \qquad \ldots\ldots(38)$$

where $C = 10^{-12}$ ergs per ion pair, r is the distance between the ions, and R_i is the ionic radius. The quantity $\rho = 0{\cdot}345$ Å. It may readily be shown that, for a face-centred cubic lattice, the contributions to $C_{11}-C_{12}$ and C_{44} are respectively:

$$\tfrac{1}{2}r^2\frac{d^2w}{dr^2} + \tfrac{7}{2}r\frac{dw}{dr} \quad \text{and} \quad \tfrac{1}{2}r^2\frac{d^2w}{dr^2} + \tfrac{3}{2}r\frac{dw}{dr}. \qquad \ldots\ldots(39)$$

The coulomb term may be calculated as described for the alkali metals.

For copper, the following comparison between calculated and experimental values may be made:

		$(C_{11}-C_{12})$			C_{44}		
	$\frac{1}{\beta}$	Coulomb contribution	Ion–ion interaction	Total	Coulomb contribution	Ion–ion interaction	Total
Calculated	14·1	0·573	4·53	5·103	2·57	6·4	8·97
Observed	13·9	—	—	5·1	—	—	8·2

All quantities are given in units of 10^{11}dyn/cm^2.

Thus the contribution from ion–ion interaction is now far more important than the coulomb contribution; for the alkali metals the opposite was true. The noble metals, however, are still markedly anisotropic, as shown by the ratio $2C_{44}/(C_{11}-C_{12}) = 3{\cdot}2$ for copper.

As for the alkali metals, the agreement between the calculations and experiment goes far to confirm the general accuracy of the model proposed for the electronic structures of the noble metals copper, silver and gold.

§ 4. THE DIVALENT METALS

In the case of the monovalent metals of groups IA and IB of the Periodic Table, the valency electrons which exist outside the closed shells of the ion are insufficient to fill the first Brillouin zone, and many of their properties are similar to those of free electrons. There is little difficulty in assessing the energies of electrons at the surface of the Fermi distribution, since the surfaces of constant energy within the first Brillouin zone approximate to spheres. Conditions are much more complicated for the divalent metals, of which we may take the elements calcium, magnesium and beryllium as typical. Calcium is face-centred cubic in crystal structure, so that the first Brillouin zone will be as illustrated in fig. 14, and able to contain a maximum of two electrons per atom. Since calcium is a truly metallic material, and conducts electricity, it is clear that there must be overlap between the first and second zones. The actual surface in k-space of maximum electronic energy will therefore be complicated, and will not approximate to a sphere. This introduces considerable complexity into the calculations. Furthermore, account must be taken, in any mathematical approach, of the interactions

between the two valency electrons which must be assumed to be contained within any one atomic polyhedron. Attempts, however, have been made to calculate the electronic condition of these metals.

Calcium, with the electronic configuration ($1s^2$; $2s^2$, $2p^6$; $3s^2$, $3p^6$; $4s^2$) in the free atom, has been studied by Manning and Krutter (1937), who applied the cellular method developed by Wigner, Seitz and Slater. The self-consistent field for the calcium atom, calculated by Hartree, was available for the basis of the calculations of the potential $V(r)$ within the atomic polyhedron. The details of the calculation appear to have followed closely those of Krutter's work on copper, and atomic s, p, d and f functions were employed for setting up the wave function for the electrons in the metal. The potential used was Hartree's core potential, and the effect of the second valency electron was included by adding the potential of the valency electron with charge density normalized to the atomic volume. Applying boundary conditions similar to those previously considered, the energy of the electronic states in the metal may then be determined as a function of **k**,

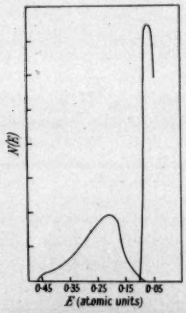


Fig. 21. Density of states plotted against energy for calcium.
(*After* Manning and Krutter 1937)

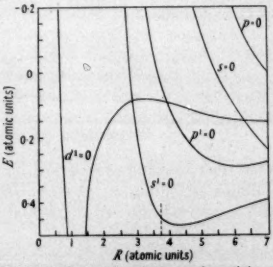


Fig. 22. Energy band system for calcium, as a function of distance between atoms.
(*After* Manning and Krutter 1937)

the wave vector; owing to the complexity of the calculations, this may be done for certain specific directions in the crystal, and Manning and Krutter choose the directions [100], [110] and [111]. Energy contours in k-space were established for chosen planes, and by a combination of graphical and numerical interpolative methods, the volume in k-space enclosed by a given energy surface was determined. From the plot of the volumes enclosed by energy contours in k-space as a function of energy, the number of states with energies lying between the energies E and $E+\Delta E$ was found by graphical differentiation, giving the distribution of states summarized in fig. 21. According to the calculations, the lowest state, corresponding to the ground state of the valency electron, varies with distance between atoms as shown by the curve $s'=0$ in fig. 22. At interatomic distances close to those actually observed for calcium, the next higher band begins at an energy on the curve which is denoted by the symbol $d'=0$, although for a greatly expanded lattice the next higher band would begin at energies along the curve denoted $p'=0$. The distribution curve of fig. 21 would thus appear to represent overlap from a first Brillouin zone containing mainly s states into a second zone in which the states

had mainly the character of d states, since the filled levels extend in energy, at the equilibrium lattice spacing, from a point on the $s^1=0$ curve to approximately the $d'=0$ curve. In the report of the work of Manning and Krutter, it is stated that the relatively low electrical conductivity is due to the small overlap illustrated in fig. 21. Under the influence of an electric field only relatively few electrons may be accelerated to higher energies, and according to the calculations the overlap can only take place at the point $(0, \pi/\sqrt{2}R, \pi/2\sqrt{2}R)$ and the twenty-three other equivalent points, where R is equal to half the interatomic distance. These points correspond with the twenty-four corners of the first Brillouin zone, so that overlap does not take place, as would be expected, in the [111] directions, in which the spherical 'free electron' Fermi surface would first touch the bounding planes of the zone, but rather in the [021] directions. This is a somewhat unexpected result, and may be partly attributable to the approximations made in the calculations, especially in connection with the interaction of the two valency electrons with each other. The calculations do, however, suggest a small overlap of electrons into the second Brillouin zone in certain specified directions only, and interpret the electrical conductivity effects. It will be noted that, in fig. 22, the curve labelled $d'=0$ is almost independent of the interatomic distance in the region where the minimum of the ground state curve occurs. Almost the whole of the change of energy with atomic separation therefore depends upon the lower electron levels, and the equilibrium lattice spacing would be expected to correspond to the minimum in the ground state curve. This interpretation, however, gives too large an estimate of the equilibrium spacing, probably owing to neglect of correlation corrections in the calculations. It would be of interest to repeat the calculations with the inclusion of appropriate corrections, since the metal calcium, according to Manning and Krutter, just misses being an insulator. No detailed calculations of the elastic constants have been made on the basis of the calculations, so that the only check with experiment lies in the approximate estimate of the lattice spacing.

The case of beryllium has been examined in considerably greater detail by Herring and Hill (1940); the calculations are very complex in nature, but do illustrate the magnitude of the problem of accurate estimates of physical properties on the basis of the electron theory. The crystal structure of beryllium is of the close-packed hexagonal type, but with an axial ratio of $1{\cdot}584_8$, instead of the ideal axial ratio of 1·633, which corresponds to the close-packing of spheres. At present the reason for the assumption, by beryllium, zinc and cadmium, of axial ratios differing widely from the ideal value is not understood. The first Brillouin zone for the hexagonal structure is illustrated in fig. 23, and can contain two electrons per atom. Since beryllium, and also the other close-packed hexagonal metals, conduct electricity, there must be overlap from the first Brillouin zone, and the nature of this overlap varies according to the axial ratio. The calculations of Jones (1934) indicate that the first complete zone is bounded by the planes $\{1\bar{1}0, 0\}$, $\{1\bar{1}0, 1\}$ and $\{000, 2\}$; there are energy discontinuities across all these planes, so that within the volume bounded by the planes of the type A, B and C, the energy varies continuously with k, whatever the direction of the vector $\mathbf{k}$. For certain directions, however, there is no energy gap along edges, such as those at which $\{1\bar{1}0, 0\}$ planes intersect, and the volume illustrated in fig. 23 is the smallest which is surrounded by an energy gap in all directions. There are thus three possibilities of overlap; there will be overlap across the vertical (internal) $\{1\bar{1}0, 0\}$ faces at less than two electrons per atom. Then overlap into the second zone may occur, at

two electrons per atom, either at the vertical intersections of the $\{1\bar{1}0, 0\}$ faces (the re-entrant corners in fig. 23) or across the horizontal $\{000, 2\}$ faces. Which of these occurs at the lower energy will depend on the axial ratio, which determines the relative distances from the origin of k-space of the places of possible overlap. For beryllium, overlap into the second zone occurs first at the re-entrant corners; at two electrons per atom the overlap across the $\{000, 2\}$ faces does not take place, but might do so if the electron/atom ratio could be raised beyond 2 by alloying. The Fermi distribution will therefore depart markedly from that characteristic of free electrons, and this is most difficult to allow for in theoretical calculations.

The calculations of Herring and Hill follow in principle the methods applied to the alkali metals and to copper, and the total energy of the crystal is estimated,

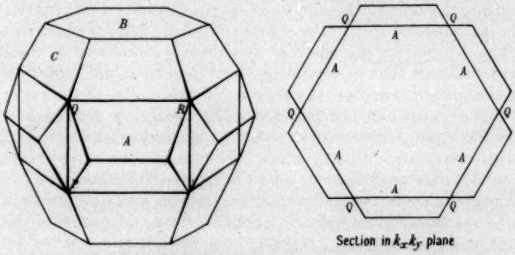


Fig. 23. First Brillouin zone for close-packed hexagonal structure.

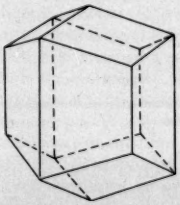


Fig. 24. Atomic polyhedron for close-packed hexagonal structure.

using for the core wave functions the 1s wave functions given by the Hartree solution for the doubly charged Be^{++} ion. Using the concept of the atomic polyhedron, which, for an hexagonal metal takes the form of fig. 24, it is shown that the potential for the valency electron inside any atom polyhedron (approximated by a sphere) is of the form

$$V(r) + \left(\frac{3X}{r_s} - \frac{Xr^2}{r_s^3}\right), \qquad \ldots\ldots(40)$$

where $V(r)$ is the potential due to the ion at the centre, and where the second term is the potential due to a charge of X electrons ($=2$ for beryllium) uniformly distributed over the volume of the sphere, which is of radius r_s.

The total energy W_H relative to widely separated Be^{2+} ions and electrons consists of three parts: (i) kinetic energy of the valency electrons, (ii) electrostatic

energy of the crystal relative to the 'decomposed' state, (iii) exchange energies, relative to those in the 'decomposed' state. Herring and Hill derive an expression for W_H in terms of the wave functions for the electron states. To obtain the binding energy for beryllium, it is necessary first to apply a correction for the correlation energy, due to the fact that electrons tend to avoid each other in a way not taken account of in the calculation of W_H. The binding energy per atom is then the difference between the resulting energy, divided by $2N$, and the energy of a neutral Be atom relative to the doubly ionized state. Difficulty arises in assessing the energy of a neutral Be atom relative to the ionized state; it is incorrect to take for this difference a value as calculated from the Hartree data (Hartree and Hartree 1935) if the correlation term referred to above is included in W_H, because the tendency for the two valency electrons to avoid each other is not taken into account in the Hartree calculations. At the same time, the correlation term must be included in W_H, so that the experimental energy difference of 2·024 Rydberg units betweenthe Be^{2+} ion and the Be atom is used by Herring and Hill.

Having derived expressions for the coulomb energy, the exchange energies, the electrostatic interaction energy of the ions, and the correlation energy, Herring and Hill were faced with the problem of determining the energies of the excited states, that is, the Fermi energy. As we have seen above, the presence of overlaps means that no treatment based upon 'free' electrons can be expected to be successful. The energy of the lowest state for the valency electron E_0 is comparatively easily obtained by the Wigner–Seitz method of solving the Schroedinger equation for an electron moving in the potential $V(r)+(3X/r_s-Xr^2/r_s^3)$, subject to the usual boundary conditions at the limits of the atomic polyhedron or sphere. Herring and Hill obtained the following values, assuming an axial ratio of 1·633 instead of the observed value of $1{\cdot}584_8$:

r_s (Bohr units)	2·07	2·37 (obs. value)	2·67
$E_0(r_s)$ (Rydberg units)	0·333	0·102	−0·017

For wave vectors lying close to zero, there will be electron energies close to E_0 whose values may be derived from the relation $E_k = E_0 + \alpha k^2$ where α is a measure of the effective mass of the electron, and may be calculated from the expression $\alpha = (m/\hbar^2)(\partial^2 E_k/\partial k^2)_{k=0}$, or developments of this equation. This treatment, however, cannot give an accurate picture of the excited energies, which account for most of the occupied quantum states, and use was therefore made of a more exact approximational method, due to Herring (1940), known as the 'orthogonalized plane wave' method. This method gives values of E_0 in good agreement with the more usual Wigner–Seitz method, but is claimed to be more accurate for the states of higher energy.

On inserting into the expression for W_H the various computed terms, the variation of W_H with r_s may be calculated; the equilibrium value of r_s in the crystal of beryllium is 2·37 atomic units and for this value the binding energy calculated, after subtracting the observed energy −2·024 Rydberg units for the energy of the neutral beryllium atom, is 0·170 Rydberg units per atom, or 53 kcal/g atom. This compares with an observed value of 85 kcal/g atom at 0° K. The discrepancy is perhaps not surprising in view of the difficulty of the work, and the use of the experimental value for the energy of the neutral atom. This procedure tends to give an underestimate of the binding energy since it automatically includes the correlation between the valency electrons but does not

compensate for the imperfections in the calculations of W_H, from which it is subtracted. Herring and Hill also suggest that the calculated energies of the excited states may be somewhat too high, which would also account for a part of the discrepancy between the observed and calculated binding energies.

The binding energies calculated for the three values 2·07, 2·37 and 2·67 atomic units of r_s may be fitted accurately on a parabola when plotted against r_s, and the minimum of this curve indicates the theoretical equilibrium value of r_s. The value deduced is 2·23 to 2·257 atomic units, which compares reasonably well with the observed equilibrium r_s value of 2·37. The compressibility, in c.g.s. units, may be expressed as

$$\frac{1}{\beta} = \frac{3{\cdot}88 \times 10^{12}}{r_s}\left[\frac{d^2}{dr_s^2}\left(\frac{W}{2N}\right) - \frac{2}{r_s}\frac{d}{dr_s}\left(\frac{W}{2N}\right)\right]. \qquad \ldots\ldots(41)$$

The second term vanishes at the theoretical equilibrium value of r_s and therefore makes a small contribution to the value of the compressibility, 0·9 to $1{\cdot}3 \times 10^{12}$ c.g.s. units calculated for $r_s = 2{\cdot}37$. Again fair agreement is obtained with the experimental value of $1{\cdot}2 \times 10^{12}$ c.g.s. units.

Having determined the energies of the excited electronic states, the relation of this energy to k may be plotted for various directions in k-space, and, hence, by

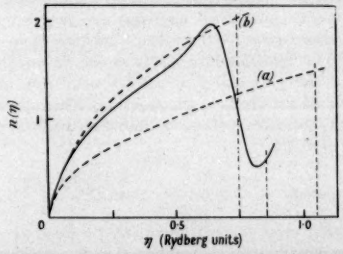


Fig. 25. Density of states plotted against energy for beryllium (full curve). The broken curves represent (*a*) free electrons, and (*b*) electrons for which $\eta_k = \eta_0 + \alpha k^2$. (*After* Herring and Hill 1940)

interpolation, contours of constant energy in k-space may be derived, allowing the number of electron states per unit energy range to be roughly determined. Figure 25 shows the resulting density of states curve; the two parabolas show the distributions which would result from the assumption of free electrons (lower curve), and from the assumption that the effective mass of the electrons is different from unity and has the value derived in the paper. The shape of the density of states curve is very similar to that experimentally observed for the K x-ray emission spectrum of beryllium by O'Bryan and Skinner (1934), but whereas the emission line would suggest a bandwidth of approximately 13·5 ev, the calculated bandwidth is only 0·872 Rydberg units or 11·8 ev. It would thus appear, from fig. 25, that the experimental bandwidth is more in agreement with the simple free electron theory than with the detailed calculations, which is a somewhat surprising result.

The density of states curve begins to rise slightly at the top of the Fermi distribution; this would be consistent with a small overlap from the first Brillouin zone. The small density of states at the highest energies is also consistent with the

magnitude of the electrical conductivity of beryllium, and the general form of the curve indicates that, like calcium, the metal has just missed being an insulator. Herring and Hill have also pointed out that the low density of states near the Fermi surface accounts for the diamagnetism of beryllium, as contrasted with the paramagnetic character of magnesium and calcium. The abnormally low density of states causes the paramagnetic contribution which the spins of the electrons make to the total susceptibility to be only a fraction of the value it would have were the electrons free.

The theoretical calculations do not, however, correctly interpret the work function of beryllium, probably because the exchange and correlation energies for the electrons near to the top of the filled region have been underestimated in the calculations.

Magnesium has also recently been studied by Raimes (1950). The Brillouin zone for magnesium is of the same form as that shown in fig. 23, but, since the axial ratio is closer to the ideal value of 1·633, the zone is less elongated in the direction of the hexagonal axis. As with beryllium, the internal overlap across the $\{1\bar{1}0, 0\}$ planes has occurred at 2 electrons per atom, followed by the overlap into the second zone at the re-entrant corners. The overlap across the $\{000, 2\}$ faces has not yet occurred, though evidence from the lattice spacings of magnesium-rich alloys indicates that a very small increase in the electron/atom ratio in sufficient to cause the onset of this overlap. The treatment adopted by Raimes follows more intimately the Wigner–Seitz approach than the work of Herring and Hill on beryllium; the metal is again considered as divided up into atomic polyhedra, which may be approximated by spheres. The ion-core field used was estimated from functions derived by Fock, assuming that the ion core remains unchanged in the solid state. By this means exchange interaction between valency and core electrons is neglected both for the metal and for the free atom, so that the error in the final cohesive energy from this cause is small. Two possibilities are available for approximating the field within which the valency electron moves. Since electrons with parallel spins are unlikely to approach closer to each other than a distance equal to the radius of an atomic cell, it may be postulated that the presence of an electron with positive spin in an atomic cell effectively excludes all others of the same spin. The required potential field in which the given electron moves is then the sum of the ion-core field and that due to the charge distribution, within the cell, of all electrons of opposite spin.

Alternatively, neglecting correlations between the positions of electrons, the potential field is the sum of the ion-core field and that due to the charge distribution within the polyhedron of all electrons. The former alternative is chosen as more physically acceptable. The potential field is thus

$$V(r) = v_c(r) + 2N \int_0 \frac{|\psi_0(r')|^2}{|r-r'|} d\tau', \qquad \ldots\ldots(42)$$

where $N|\psi_0(r)|^2$ represents the charge density, due to electrons with one type of spin only, at the point r; N is the number of atoms in the crystal. The wave function for the lowest state of the electron may now be obtained by solving the Schroedinger equation subject to the usual boundary conditions. It will be noted that since $V(r)$ depends upon the wave function, the solution requires a method of successive approximation, and the method is laborious. The type of wave function obtained, however, is illustrated in fig. 26, which shows the function calculated for

$r_s = 3{\cdot}28$ atomic units (observed equilibrium value $= 3{\cdot}35$ atomic units). It should be realized that a considerable amount of computation is necessary; whereas for the alkali metals a single integration will provide two points on the curve of E_0 against r_s, in this case it is necessary, since $V(r)$ is a function of ψ, to fix r_s arbitrarily, choose E_0 to satisfy the boundary conditions, and recalculate $V(r)$ several times before obtaining a single point on the curve. Solutions were, however, obtained covering the most important range of r_s, and the ground state curve is shown in fig. 27, E_0.

The metal may now be regarded as $2N$ valency electrons moving in a lattice of N doubly charged positive ions, and the total energy of the crystal may be written

$$E_t = 2N(E_0 + E_F + E_x + E_c) \qquad \ldots\ldots(43)$$

where E_F is the Fermi energy per electron, E_x is the exchange energy due to electrons of parallel spin keeping apart, and E_c is the correlation energy due to

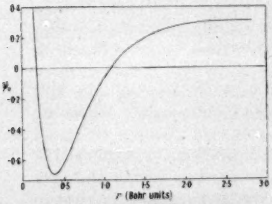


Fig. 26. Wave function of electron in lowest state for magnesium (calculated for $r_s = 3{\cdot}28$ atomic units).
(*After* Raimes 1950)

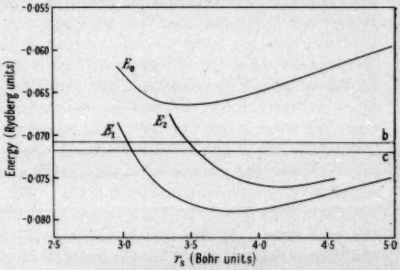


Fig. 27. Curves, against r_s, of E_0, E_1 and E_2 for magnesium. E_1 represents total energy assuming effectively free electrons; E_2 represents total energy after correction for the effective mass of the electron. Lines b and c denote respectively the mean energy of a valency electron in the free atom without and with correlation correction.
(*After* Raimes 1950)

electrons with antiparallel spins. Expressions are derived for these various terms, and the total energy per electron E_1 is given as:

$$E_1 = E_0 + E_F + E_x + E_c = E_0 + \frac{3{\cdot}51}{r^2} - \frac{1{\cdot}154}{r} - \frac{0{\cdot}576}{0{\cdot}79r + 51}. \qquad \ldots\ldots(44)$$

The curve of E_1 against r_s is also shown in fig. 27. As in the case of beryllium, to calculate the binding energy the mean energy of the valency electron in the free atom E_A must be known for subtraction from E_1. Because the experimental value for E_A automatically includes the exchange energy between core and valency electrons and other effects not included in the calculation of E_1, and also because the imperfections of the ion-core field used in the calculations are not balanced by corresponding faults in E_A, Raimes considered it preferable to calculate E_A, making similar assumptions to those made in calculating E_1. Making reasonable estimates for the energy contribution due to correlations between the positions of the two valency electrons in the free atom, the value $E_A = -0{\cdot}719$ Rydberg units per electron was derived. The binding energy (heat of sublimation) thus becomes

the difference between E_A and the minimum in E_1, or 43 kcal/g atom as compared with the experimental value of 38 kcal/g atom.

The value of r_s at the minimum in E_1 corresponds to 2·0 Å., compared with the experimental value of 1·77 Å. The theoretical compressibility, which is calculated from the expression

$$\frac{1}{\beta} = \frac{1}{6\pi\rho}\left(\frac{d^2E_1}{dr^2}\right)_{r=\rho}, \qquad \ldots\ldots(45)$$

where ρ is the equilibrium value of r_s, is $3{\cdot}2\ \mathrm{cm^2/kg} \times 10^{-6}$ as compared with the observed value of $3{\cdot}0\ \mathrm{cm^2/kg} \times 10^{-6}$.

The agreement between the calculated and observed quantities is good, considering the approximations made, which may be examined somewhat more closely.

(i) Fermi energy. For the derivation of the values quoted above, it was assumed that the valency electrons were effectively free. In view of the fact that magnesium must involve an overlap of electrons from the first Brillouin zone, the agreement between theoretical and observed values quoted above is surprising. In examining this point, Raimes applied a correction by making an estimate of the effective mass of the electron, as was done by Herring and Hill for beryllium. This correction shifts the curve for the total energy of the crystal per electron to higher energies, thus increasing the calculated value of r_s to 2·2 Å., and decreasing the cohesive energy to 25 kcal/g atom. The compressibility is less affected, and takes the value $3{\cdot}4\ \mathrm{cm^2/kg} \times 10^{-6}$. Agreement with experiment thus deteriorates as a result of the correction, and it would appear that the approximation of free electrons is not a bad approximation for magnesium. The energy gaps at the zone boundaries appear to distort the distribution of states curve from the parabolic form in such a way that the resulting Fermi energy differs little from that of free electrons.

(ii) Interactions between valency electrons and core electrons. These interactions were neglected; this was considered justifiable because the ion of magnesium is small compared with the atomic cell. Further, since the calculation for the free atom also neglected this factor, the error in the energy of the metal may be expected to be small.

(iii) Exchange and correlation energy calculations. These were evaluated using the approximation of free electrons, which is not true for real metals, but it is estimated that the effects of this procedure are small. Thus the possible error of 20% which is usually accepted in the calculation of the correlation energy for free electrons causes an uncertainty of only 0·01 Rydberg unit per electron (6 kcal/mole) in the value of the cohesive energy. The correlation correction in the calculations for the free atom was, however, made by an empirical method, and may be somewhat in error.

It would appear, therefore, that the major uncertainty in the final calculation lies in the necessary neglect of the effect of the Brillouin zone boundaries, which, in this case, seems to operate in such a way as to make the assumption of free electrons accidentally a good one. Although the curve for the density of states cannot be derived from the present work, if the electrons were really free the bandwidth would be approximately 7·3 ev, whereas the bandwidth corresponding to the soft x-ray emission band of magnesium is only $6{\cdot}2 \pm 0{\cdot}3$ ev.

As noted above, in magnesium there is overlap of electrons from the first Brillouin zone in directions at right angles to the hexagonal axis. Overlaps in the

direction of the hexagonal axis may be set up if the number of electrons per atom is increased by alloying. It has been shown by Jones (1934) that a small overlap of electrons across the boundary planes of a Brillouin zone causes an internal stress in the crystal. In fig. 28, O represents the origin of k-space, and PQ is a section of a plane of energy discontinuity, perpendicular to OA. Curve S represents a section of a surface of constant energy. The total energy per atom corresponding to the states between S and PQ is

$$U = 2\tau \int E(k) dv, \qquad \ldots\ldots(46)$$

where dv is an element of volume in k-space and τ is the atomic volume, the density of states in k-space being 2 per unit volume of the crystal. The value of k at A is inversely proportional to the lattice spacing in the corresponding direction in the crystal, so that, if the crystal undergoes a strain $\delta\epsilon$, the component of k parallel to

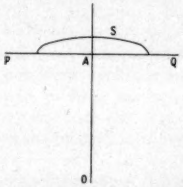


Fig. 28. Overlap of electrons across plane PQ in k-space.
(*After* Jones 1934)

OA changes by an amount $-k\delta\epsilon$, and the surface S will change slightly. Assuming that the number n of states per atom enclosed between S and PQ remains constant during the strain, then

$$\frac{\partial U}{\partial \epsilon} = 2\tau \int \frac{\partial E}{\partial \epsilon} dv. \qquad \ldots\ldots(47)$$

If σ is the stress (dyn/cm^2) which produces the strain ϵ, then

$$\tau\sigma = -n\partial \overline{E}/\partial\epsilon, \qquad \ldots\ldots(48)$$

where $\partial \overline{E}/\partial\epsilon$ is the average value of $\partial E/\partial\epsilon$ taken over the n states between S and PQ. The value of $\partial \overline{E}/\partial\epsilon$ may be approximated by considering the overlapping electrons as free; when the component of k parallel to OA changes by $-k\delta\epsilon$ as a consequence of a crystal strain $\delta\epsilon$ we may write

$$dE/d\epsilon = -k\, dE/dk = -2E \qquad \ldots\ldots(49)$$

since for free electrons $E = h^2k^2/2m$. Thus, the stress produced by n electrons lying in a small region beyond a plane of energy discontinuity in k-space acts in a direction perpendicular to the plane, and is given approximately by the expression

$$\sigma = 2n\overline{E}/\tau \qquad \ldots\ldots(50)$$

where $\overline{E}$ is the mean value of the energy. If a small change in composition of an alloy causes a change in the number of overlapping electrons from n to $n+\delta n$, then the increase in the relevant component of the internal stress is given by

$$\delta\sigma = 2\overline{E}\,\delta n/\tau. \qquad \ldots\ldots(51)$$

Thus an increase in the number of overlapping electrons tends to expand the lattice, and a corresponding decrease to contract the lattice, in a direction at right angles to the overlapped plane. This theory has been applied by Jones

(1934) with success to the variation in the axial ratios of the close-packed hexagonal ϵ-phase in the copper–zinc system, and the close-packed hexagonal solid solution of copper in zinc. The lattice spacings of magnesium alloys are also interpretable in terms of the internal stress due to overlapping electrons, as shown by Raynor (1940) and in subsequent papers (Hume-Rothery and Raynor 1940, Raynor 1942). The solution, for example, of indium in magnesium causes an increase in the electron/atom ratio, and the lattice spacing is contracted at right angles to the hexagonal axis much less than when cadmium is dissolved in magnesium, in spite of the fact that the atomic sizes of indium and cadmium are similar. This is due to the increase in the number of electrons overlapping from the first Brillouin zone in directions at right angles to the hexagonal axis, which causes an expansion relative to the distortion produced by cadmium, which leaves the overlap almost unaffected. Similarly, the addition of one atomic per cent of tin increases the overlap to a higher degree than one atomic per cent of indium, and the resultant contraction of the lattice is less than that produced by indium. By contrast, the monovalent metal silver produces a reduction in the overlap, and hence a marked contraction relative to that produced by cadmium. The lattice spacings of some alloys of magnesium are illustrated in fig. 29.

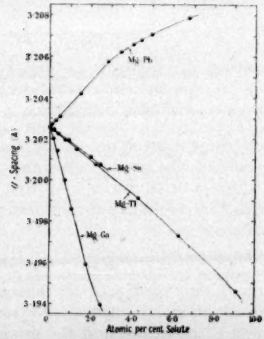


Fig. 29. (a-spacing, composition) curves for magnesium-rich alloys.

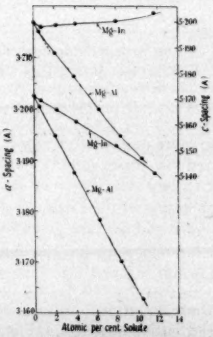


Fig. 30. (a- and c-spacing, composition) curves for magnesium-rich alloys.

The c spacings (parallel to the hexagonal axis) show more complex relationships, when plotted in terms of alloy composition. The curve for cadmium in magnesium is smooth and continuous, showing a uniform contraction, and the axial ratio c/a remains almost unchanged. On solution of indium, the c spacing at first contracts, and then, at 0·75 atomic per cent of indium, begins to rise sharply. This is due to the onset of the overlap across the {000,2} faces of the first Brillouin zone. When this takes place, there is an internal stress in the crystal, acting parallel to the hexagonal axis, which is non-existent in the crystal of magnesium, in the magnesium-rich magnesium–cadmium alloys, and in the magnesium–indium alloys below 0·75 atomic per cent of indium. The result of this stress is to tend to expand the c spacing of the alloys, and since the resistance to expansion

or contraction parallel to the hexagonal axis is relatively less than that in the perpendicular direction, the resulting effect is large. Above 0·75 atomic per cent of indium, therefore, the c spacing is expanded by the onset of, and subsequent increase in, the overlap across the {000,2} faces of the zone. This effect outweighs the atomic size effect, which would lead to a continued contraction of the type exhibited below 0·75 atomic per cent of indium. It has been shown that a similar effect occurs at the same electron/atom ratio of 2·0075 for other trivalent solute metals, and also for the quadrivalent metals tin and lead. Silver, which decreases the electron/atom ratio, shows no such effect. The (c-spacing, composition) curves for various magnesium alloys are shown in fig. 30.

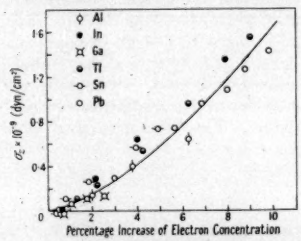


Fig. 31. σ_z plotted against percentage increase in electron/atom ratio.
(*After* Jones 1949)

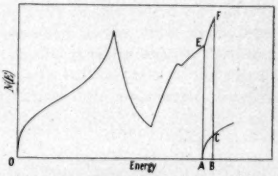


Fig. 32. Diagrammatic representation of overlap across {000,2} planes in k-space for magnesium.

The more quantitative aspect of the electron-overlaps in magnesium has been examined by Jones (1949), who has considered the pure magnesium crystal subject to two stresses: (i) a hydrostatic pressure or tension σ, and (ii) a stress in the direction of the c axis, σ_z. The two together result in a shear stress. The strains $\delta a/a$ and $\delta c/c$ are given by the equations

$$\delta a/a = (S_{11}+S_{12})\sigma + S_{13}\sigma_z, \qquad \delta c/c = 2S_{13}\sigma + S_{33}\sigma_z \qquad \ldots\ldots(52)$$

where S_{11}, S_{12}, S_{13} and S_{33} are the appropriate elastic coefficients for magnesium. From these equations σ and σ_z at any composition may be determined from the measured lattice spacings. Figure 31 shows σ_z plotted against the percentage increase in the electron/atom ratio (experimental points); since all the solutes are not of the same valency, it is clear that σ_z is determined by the concentration of electrons, and not of atoms. If x is the number of added electrons per atom of alloy, then the overlap of electrons across the {000,2} faces of the zone may be assumed to occur for x a little greater than zero. Referring now to fig. 32, the area OEA represents approximately $2L$ electrons, where L is the number of atoms. On increasing the number of electrons slightly, the point B is reached, and the area AEFB, which may be taken as a rectangle for the present purpose, corresponds to xL electrons where x is the number of electrons added. The number of states per energy range E to $E+dE$ may be taken as a constant N between A and B, so that $N\epsilon = Lx$, or $x \propto \epsilon$, where ϵ is the difference in energy between points A and B. The contribution of the overlapping electrons at the {000,2} faces to the area AEFB is given by the area of the parabolic curve ACB,

since at the bottom of a zone $n(\epsilon)=\epsilon^{1/2}$ where $n(\epsilon)$ is the number of overlapping states per unit energy range, assuming the energy at A to be zero. The area ACB is thus

$$\int_0^\epsilon n(\epsilon)\,d\epsilon=\int_0^\epsilon \epsilon^{1/2}\,d\epsilon=\tfrac{2}{3}\epsilon^{3/2}. \qquad \text{.....(53)}$$

Thus the number of overlapping electrons is proportional to $\epsilon^{3/2}$ and hence to $x^{3/2}$. If the c spacings of magnesium alloys, therefore, are controlled mainly by the electronic overlap, σ_z should be proportional to $x^{3/2}$. The full curve of fig. 31 is given by the equation $\sigma_z=\text{const.}\,x^{3/2}$, with the appropriate choice of constant, and it may be seen that the theory is in good agreement with the facts.

A considerably more detailed calculation of these effects has recently been given by Jones (1950), taking into account not only the overlapping electrons but the holes remaining in the first zone, and the calculations have been extended to cover the case of decreasing electron/atom ratio. A parameter S is chosen such that $S=(c/a)^{2/3}$, so that $a=(4\tau/\sqrt{3})^{1/3}S^{-1/2}$ and $c=(4\tau/\sqrt{3})^{1/3}S$, where τ is the atomic volume.

It is then assumed that changes in volume arise from forces acting equally in all directions (i.e. a uniform pressure or tension) which produces a shearing strain (in an anisotropic elastic crystal) given by

$$\frac{S-S_0}{S_0}=\tfrac{2}{3}k\left(\frac{\tau-\tau_0}{\tau_0}\right), \qquad \text{.....(54)}$$

where k depends on the elastic coefficients of the crystal. If this strain is subtracted from that observed, then the resultant is the strain induced by the solute atoms at constant volume, which can thus be determined from the experimental results. This quantity may then be compared with the relation between S and n, where n is the increase of the number of electrons per atom in the alloy, which is derived from the equilibrium condition:

$$\left(\frac{\partial W(\tau,s,x)}{\partial s}\right)_{\tau,x}=0 \qquad \text{.....(55)}$$

where W is the energy per atom, expressed relatively to the separated atoms, and x is the concentration of solute. Thus, if $\delta a/a$ and $\delta c/c$ are the observed strains, then for the change of S at constant volume,

$$\left(\frac{S-S_0}{S_0}\right)_\tau=\frac{2}{3}\left(\frac{\delta c}{c}-\frac{\delta a}{a}\right)-\frac{2}{3}k\left(\frac{\delta c}{c}+2\frac{\delta a}{a}\right). \qquad \text{.....(56)}$$

This process eliminates the effect of the different atomic sizes in the alloys. The variations of S at constant volume depend only on n, while the volume may now no longer be treated as a variable, so that the condition for equilibrium becomes

$$\left(\frac{\partial W(s,n)}{\partial s}\right)_n=0. \qquad \text{.....(57)}$$

Making the reasonable assumption that the Fermi energy is the only part of W which need be considered, and considering the overlaps already referred to together with the remaining holes in the corners of the first Brillouin zone, it can now be shown that

$$\left(\frac{S-S_0}{S_0}\right)_\tau=A(n-n_0)^{3/2}+Bn+Cn^2 \qquad \text{.....(58)}$$

where the symbols have the following meanings

$$A=\frac{4}{3}\left(\frac{\zeta}{\mu\tau}\right)\frac{\alpha}{N^{3/2}} \qquad \text{if } n>n_0 \quad \text{or } 0 \quad \text{if } n<n_0$$

$$B=\frac{2}{N}\left(\frac{39}{41}N_D-N_A\right)\left(\frac{\zeta}{\mu\tau}\right)$$

and

$$C=\frac{1}{N^2}\left(\frac{39}{41}\frac{dN_D}{d\zeta}-\frac{dN_A}{d\zeta}\right)\left(\frac{\zeta}{\mu\tau}\right).$$

ζ is the energy of the top of the Fermi distribution, while the density of states of single overlap regions is perpendicular to the hexagonal axis. N_D is the density of vacant states in the zone corners and N is the total density of states in magnesium. The symbol μ denotes the sum of elastic constants

$$(C_{11}+C_{12}+2C_{33}-4C_{13})$$

and α is a constant. By estimating N_A, N_B and N_D, and assuming 7·3 ev for ζ, Jones shows that, for $n>n_0$,

$$[(S-S_0)/S_0]_\tau=-0{\cdot}213n^2+0{\cdot}150(n-0{\cdot}005)^{3/2} \qquad \ldots\ldots(59)$$

and for $n<n_0$,

$$[(S-S_0)/S_0]_\tau=-0{\cdot}213n^2. \qquad \ldots\ldots(60)$$

This result gives the full curve in fig. 33, and the experimental points are plotted for comparison. Agreement is good except for silver in magnesium.

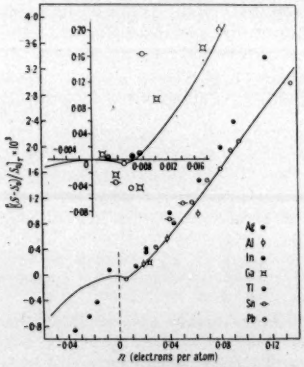


Fig. 33. The quantity $[(S-S_0)/S_0]_\tau$ plotted against change in number of electrons per atom for magnesium-rich alloys.
(*After* Jones 1950)

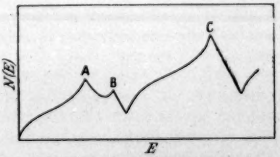


Fig. 34. Density of states plotted against energy for the face-centred cubic structure. The peaks A, B, and C arise respectively from the {111} {200} and {220} planes in k-space.
(*After* Mott and Jones)

It may thus be seen that we have a considerable knowledge of the electronic structure of magnesium. Although there is some doubt as to the final accuracy of the calculation of the cohesive energy, the quantities derivable are of the correct magnitude, while the treatment of the lattice spacing variations on the basis of zone overlaps leaves little doubt that the scheme of overlaps described is essentially correct. Magnesium is thus of considerable theoretical interest because, in the pure metal, the overlap across the {000,2} planes of the Brillouin zone has not

occurred; this overlap comes into play, however, when the electron/atom ratio rises above 2·0075. The electronic structure of magnesium has considerable influence on its alloying behaviour. For details of this, and of the probable reason for the differences between the alloying behaviour of magnesium on the one hand and of zinc and cadmium on the other, reference may be made to the work of Hume-Rothery and Raynor (1938).

No detailed treatment has been made of the electronic structures of the B sub-group metals zinc, cadmium and mercury. Zinc and cadmium have close-packed hexagonal structures, with large axial ratios. The result of this is that overlap from the first Brillouin zone takes place first from the top and bottom of the zone, across the {000,2} faces. Apart from this, little can be said of these metals. The evidence which may be drawn from the x-ray emission spectra of zinc is referred to later with reference to the alloys of zinc with copper and with nickel.

§ 5. ALUMINIUM

Aluminium is a trivalent metal with a face-centred cubic crystal structure. The first Brillouin zone is therefore of the same form as that for copper and silver, and can contain, if completely filled, two electrons per atom. Since there are three valency electrons to be accommodated (two 3s electrons and one 3p), it is clear that there must be considerable overlap into the second zone. This might be expected to occur across the hexagonal {111} faces, or across both the hexagonal {111} and square {200} faces. The difficulty of calculating the energies of the electrons in the second zone, the form of the energy contours, and the distribution of energy levels is considerable, and until recently it was necessary to assume that the density of states was adequately represented by the curve suggested by Mott and Jones for the face-centred structure of copper (fig. 34), filled up to a level corresponding to 3 electrons per atom. There was, however, no guarantee that the energy gaps in the aluminium case remained comparable with those in copper, and the use of these curves was therefore a very approximate procedure.

The electronic structure of aluminium has now been investigated in detail by Matyas (1948), using the general methods of the Bloch approximation of tight binding. In essence, the ground state of the atom is taken as an s state, and the next p state is considered to have almost the same energy, so that the zero-order wave function for the lattice will contain both s and p functions. There are therefore to be considered the spherically symmetrical wave function belonging to the s state, $\phi_0(r)$, together with three p wave functions $\phi_1 = xf(r)$, $\phi_2 = yf(r)$ and $\phi_3 = zf(r)$. The wave functions for electrons in the lattice can be expressed as

$$\psi_{nk} = \sum_l e^{ia(kl)} \phi_n(r - al) \qquad \ldots\ldots(61)$$

where $n = 0, 1, 2, 3 \ldots$, a is the lattice constant, and al gives the position of any atom, the components of l being integers. The problem is to find how the functions ψ_{nk} derived from s and p states combine to form the true zero order wave functions for the electron in the lattice. The mathematical handling of this problem is complex and beyond the scope of the present review, but it may be shown that the energies of the valency electrons in the s and the three p bands are derivable as functions of k. Solving for the [111] direction as the direction of propagation,

$$E^{I,II} = -\tfrac{1}{2}(M+N) + 2(A_1' - 3A_0')\cos^2 \tfrac{1}{2}k - 4A \sin^2 \tfrac{1}{2}ak$$
$$\pm \{[\tfrac{1}{2}(M-N) + 2(A_1' + 3A_0')\cos^2 \tfrac{1}{2}ak - 4A \sin^2 \tfrac{1}{2}ak]^2 + 48l^2 \sin^2 ak\}^{1/2} \quad ..(62)$$

where the A's are values of integrals arising in the mathematics, and where $-M=E_0-A_0$ and $-N=-E_1-A_1$. E_0 and E_1 are respectively the energies of s and p states in the isolated atom. E^{I} refers to the energy spectrum inside the first Brillouin zone and E^{II} to the second Brillouin zone. Putting in values for the various integral expressions, the energy contours in the first two zones may be represented as in fig. 35, which includes also the results for [100] directions. Figure 35 (*a*) summarizes the contours in the xy plane, while fig. 35 (*b*) shows a section containing the [111] direction and the z axis. The energy values for the points denoted by small letters in the diagrams are as follows: e (just inside first zone), $-M$; e (just inside second zone), $-N-8A$; t, $-N+4A_1'$; b (just inside first zone), $-M+4A_0'$; b (just inside second zone), $-N-12A_1'$; c, $-N-4A_1'$. The expected shapes of the energy contours near the faces of the first Brillouin zone were examined, and it was possible to show that whereas inside the first zone the contours near to one of the hexagonal faces approximated to hyperboloids, the surfaces outside the same faces (i.e. in the second zone) were ellipsoidal in form. In the neighbourhood of the square faces, energy surfaces in the first

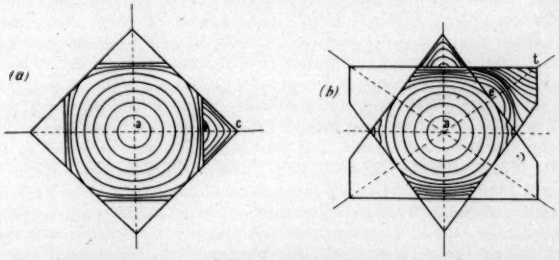


Fig. 35. Contours of constant energy for first and second Brillouin zones for aluminium: (*a*) section through reciprocal space at $k_z=0$; (*b*) section through reciprocal space containing axis k_z and direction [111].
(*After* Matyas 1948)

zone are practically parallel to the faces themselves. This gives rise to a high density of states for the energy range corresponding to the surfaces nearly parallel to these faces.

After deriving the electronic structure of aluminium by the Bloch method assuming tight binding, Matyas considered the problem in terms of the approximation of nearly free electrons; this treatment gives results consistent with the hyperbolic and ellipsoidal form of the energy contours in the first and second zones in the region of {111} faces of the first Brillouin zone, but gives little indication of the precise form of the contours just inside the {200} faces.

The distributions of electron energies which may be derived from Matyas' treatment are as shown in fig. 36. Electrons at the bottom of the first zone behave as if effectively free, and the energy surfaces are spherical; marked departure from spherical form occurs at higher energies. The energy of free electrons corresponding to the mid-point of the hexagonal faces is easily calculated to be $E_0'=\hbar^2(\sqrt{3}\pi/a)^2/2m=6{\cdot}5$ ev. The electronic energy near the mid-point of the square faces may be calculated to be approximately 7·2 ev greater than this, leading to a value of some 13·7 ev. The energy gaps across the faces {111} and

{200} are respectively $\Delta E_{111} = 3{\cdot}4$ ev and $\Delta E_{200} = 3{\cdot}1$ ev, as calculated from the atomic structure factors and structure amplitudes for the planes in the crystals. Thus, points just inside the second zone near the mid-points of hexagonal faces have energies of $6{\cdot}5 + 3{\cdot}4 = 9{\cdot}9$ ev, which is less than the energy of point b in the first zone. Overlap of electrons therefore occurs in the [111] direction. In order to find the Fermi energy corresponding to the highest occupied states, it is assumed that the first zone is nearly full and that one electron per atom lies in the second zone. Assuming the energy contours to be ellipsoids, Matyas estimates that the Fermi surface is approximately 3·5 ev higher than the state just inside the second zone near the mid-point of a hexagonal face, i.e. approximately 13·4 ev. It would therefore not be expected that any overlap at all would take place across the {200} faces of the zone, but rather that small holes, parallel to the {200} faces, would exist in the first zone. The breadth of the energy spectrum is of the order of 13·4 ev, which is in fair agreement with the experimentally derived breadth of approximately 12 ev given by the soft x-ray emission spectra experiments of Skinner (1940). The L_{III} emission band for aluminium is shown in fig. 37, and can be shown to be in reasonable agreement with the distribution of energy levels in fig. 36 when due allowance is made for the effect of transition probabilities. In connection with this, it should be noted that the wave functions for all the

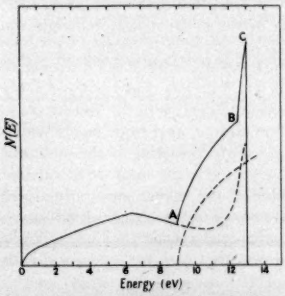


Fig. 36. Density of states plotted against energy for aluminium. The sum-curve ABC represents contributions from both first and second zones.
(*After* Matyas 1948)

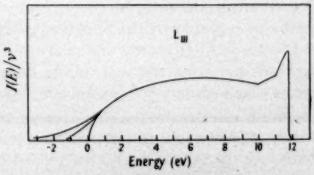


Fig. 37. Diagrammatic representation of L_{III} emission band for aluminium.
(*After* Skinner 1940

states of energy in the first zone approximate to s wave functions, with a small amount of p-type character introduced near the zone boundaries. Towards the top of the occupied levels, the contribution from p-type wave functions is considerable. Thus, since at high energies the density of states is very high, we may expect an intense emission, as observed. Further, there is a peak in the distribution of states at the energy value corresponding to contact of the energy contours with the {111} planes (6·5 ev). Since these states are mainly s states, a maximum in the emission band near this energy value is also expected, and reference to fig. 37 shows that such a maximum is observed at 6·6 ev, in good agreement with theory.

According to Matyas[1] calculations, therefore, the electronic structure of aluminium is such that the first zone is nearly full, but contains unoccupied

regions in the form of holes bounded by the {200} faces and energy contours nearly parallel to these faces (fig. 35). Overlap occurs across the {111} faces, and approximately one electron per atom occupies states in the second zone. The apparent absence of any overlap across the {200} faces at three electrons per atom is somewhat surprising.

There is, however, evidence that the electronic structure of aluminium is somewhat different from that deduced by Matyas. Leigh (1951 a), in a discussion of the elastic constants of aluminium, has shown that it is necessary also to take account of an overlap of electrons across the square {200} faces of the zone. As shown previously, the change in energy of a metal on distortion is due partly to the valency electrons and to their electrostatic interaction with the ions, and partly to non-coulomb forces between the ion cores. In aluminium, the latter contribution is negligible because of the relatively large distance between ions. The electrostatic interaction, for cases where the Fermi surface lies entirely within the first Brillouin zone, gives rise to a marked anisotropy in the elastic constants, and the effect on the Fermi energy of movements of the bounding planes of the zone as the metal is sheared is small. Aluminium, however, is almost isotropic, so that the zone structure must be responsible for the difference between the behaviour of this metal and the monovalent metals. The actual elastic constants are $(C_{11}-C_{12})/2 = 2{\cdot}085 \times 10^{11}$ dyn/cm^2 and $C_{44} = 2{\cdot}853 \times 10^{11}$ dyn/cm^2; the former refers to shear across the (110) planes in the $[1\bar{1}0]$ direction, while the latter refers to shear across (100) planes.

It may be shown that, as for magnesium, the wave function of the lowest state is very flat near the boundaries of the atomic cell. There is thus no appreciable contribution to the elastic constants form the kinetic energy of the lowest state. Contributions do arise, however, from the Fermi energy, and from terms which represent respectively the potential energy of a system consisting of the ion cores and a charge distribution of density $-Zeu_0^2(r)$ where Z is the number of valency electrons per atom and $u_0(r)$ is the wave function of the lowest state normalized over a single atomic cell, and the self potential energy of the charge distribution $-e[Zu_0^2(r)-\rho(r)]$, where $\rho(r)$ is the density of valency electrons*. The coulomb interactions arising from these two terms are shown by Leigh to correspond with the following values of the shear constants:

$$(C_{11}-C_{12})/2 = Z_{\text{eff}}{}^2\, 0{\cdot}183 \times 10^{11}\ \text{dyn/cm}^2, \qquad C_{44} = Z_{\text{eff}}{}^2\, 1{\cdot}639 \times 10^{11}\ \text{dyn/cm}^2,$$

where $Z_{\text{eff}}{}^2$ is a factor related to the number of valency electrons per atom, and may be shown to have a probable value of 6·97. Thus $(C_{11}-C_{12})/2 = 1{\cdot}3 \times 10^{11}$ dyn/cm^2 while $C_{44} = 11{\cdot}4 \times 10^{11}$ dyn/cm^2, which is much greater than the value observed.

The Fermi energy therefore is very important in controlling the elastic constants. and the problem arises of dealing with the complex nature of the energy surfaces when the electrons have overlapped from the first Brillouin zone. It is possible to consider separately first the contributions from a completely filled first zone, and then the effect of the overlapping electrons. When the metal lattice is sheared, the zone is of course modified, and the altered positions of the bounding planes, and of the corners where they meet, may be determined. From this, it is possible to calculate the contributions to the strain energy for shears corresponding to $(C_{11}-C_{12})/2$ and C_{44}, subject to the approximation that

* $\rho(r) = 2\Sigma_k |\psi_k(r)|^2$

the actual energy surfaces for the first Brillouin zone may be regarded as spherical surfaces given by the relation $E(k)=\alpha_0 h^2k^2/8\pi^2m$. The resulting contributions are: $(C_{11}-C_{12})/2=\alpha_0 1{\cdot}5\times10^{11}$ dyn/cm^2, $C_{44}=\alpha_0 4{\cdot}5\times10^{11}$ dyn/cm^2. α_0 for aluminium is not known, but cannot be far different from unity.

For assessing the contribution of the overlapping electrons, it is assumed that during shear, the energy surfaces over a face of the zone move with that face and do not change relatively to it. Then, as the distances of the faces from the origin of k-space vary consequent upon shear, two effects may be postulated: the energy surfaces as a whole shift with the faces, and electrons will transfer from those faces receding from the origin to those approaching it. Developing this quantitatively, Leigh shows that the contributions to the elastic constants per atom from overlapping electrons over the hexagonal and square faces are:

	Hexagonal faces	Square faces
$(C_{11}-C_{12})/2$	$8n_aE_a/3$	$2n_bE_b-2N_bE_b^2$
C_{44}	$8n_aE_a/3-16N_aE_a^2/9$	$2n_bE_b$

n_a and n_b are the numbers of electrons per atom in the equilibrium position over one pair of hexagonal and square faces respectively, N_a and N_b are the corresponding contributions to the density of states per atom at the Fermi surface, and E_a and E_b are the equilibrium position energies at the centres of the faces.

It will be observed that several unknown parameters appear in the results, so that direct calculation of the elastic constants is not possible. It is, however, possible, by assuming reasonable values for these parameters, to account for the observed constants in terms of the three contributions discussed above. If the energy surfaces in the second zone are approximated to spheroids, it may be shown that:

$$n_a=\frac{\Omega}{4\pi^3}\frac{4\pi}{3}\frac{1}{(\alpha_a)^{1/2}}\left[\frac{8\pi^2m}{h^2}(\zeta_0-E_a)\right]^{3/2} \qquad \ldots\ldots(63)$$

and

$$N_a=3n_a/2(\zeta_0-E_a) \qquad \ldots\ldots(64)$$

with similar expressions for n_b and N_b. Since ζ_0 is known to be 11·8 ev from the soft x-ray emission spectra observed by Skinner (1940), and since the density of states per atom $N(\zeta_0)$ at the Fermi surface is also known from electronic specific heat data (Kok and Keesom 1937), it is possible to determine n_a, n_b, N_a and N_b in terms of assumed or estimated values of α_a and α_b. Choosing appropriate parameter values to reproduce the observed values of the elastic constants, the results may be summarized as follows:

Contributions	$(C_{11}-C_{12})/2$	C_{44}
Coulomb terms	$1{\cdot}3\times10^{11}$ dyn/cm^2	$11{\cdot}4\times10^{11}$ dyn/cm^2
First zone	$1{\cdot}5\times10^{11}$ dyn/cm^2	$4{\cdot}5\times10^{11}$ dyn/cm^2
Overlapping electrons	$-0{\cdot}7\times10^{11}$ dyn/cm^2	$-13{\cdot}05\times10^{11}$ dyn/cm^2
Total	$2{\cdot}1\times10^{11}$ dyn/cm^2	$2{\cdot}85\times10^{11}$ dyn/cm^2

$Z_{\text{eff}}^2=6{\cdot}97$, $E_a=9{\cdot}08$ ev, $\alpha_a=1{\cdot}94$, $E_b=11{\cdot}21$ ev, $\alpha_b=10{\cdot}21$, $n_a=0{\cdot}242$ per atom, $n_b=0{\cdot}011$ per atom, $N_a=0{\cdot}1334$ per ev per atom, $N_b=0{\cdot}027$ per ev per atom.

It will be seen, therefore, that the large contribution of the coulomb terms to the C_{44} shear constant is compensated by the effect of overlapping electrons, to which the almost isotropic nature of aluminium is mainly due.

It is of interest to note that if overlap occurred only over the hexagonal faces of the first zone, $4n_a$ would be equal to unity and $4N_a$ would be equal to $N(\zeta_0)$,

and the contributions from overlapping electrons to $(C_{11}-C_{12})/2$ and C_{44} are respectively $2E_a/3$ and $2E_a/3-4N(\zeta_0)E_a^2/9$. Since $N_a=3n_a/2(\zeta_0-E_a)$, $\zeta_0-E_a=2{\cdot}4$ ev and $E_a=9{\cdot}4$ ev, and the appropriate contributions to $(C_{11}-C_{12})/2$ and C_{44} are $6{\cdot}1\times10^{11}$ dyn/cm^2 (6·3 ev per atom) and $-17{\cdot}3\times10^{11}$ dyn/cm^2 ($-17{\cdot}8$ ev per atom). The very large negative contribution to C_{44} clearly overcompensates for the coulomb term, while the contribution to $(C_{11}-C_{12})/2$ is itself larger than the observed value for this constant. This value can only be reduced by postulating overlap across the square faces of the zone, and this analysis lends considerable support to such an electronic structure for aluminium. The contributions to the density of states from overlapping electrons are summarized in fig. 38; curve (*a*) refers to the hexagonal faces and curve (*b*) to the square faces. The full curve is the sum of (*a*) and (*b*). The point at which there is just no overlap across the square faces corresponds with 2·67 electrons per atom.

Since the contributions of the overlapping electrons to the elastic constants have been determined, it is possible to estimate the change in elastic constants expected on alloying aluminium with a metal of higher or lower valency, assuming

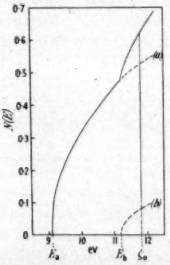


Fig. 38. Density of states plotted against energy for overlapping electrons in aluminium. Curve *a* corresponds to overlap across {111} planes, and curve *b* to the small {200} overlap.
(*After* Leigh 1951 a)

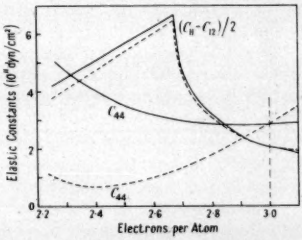


Fig. 39. Variation of elastic constants with electron/atom ratio for aluminium alloys.
(*After* Leigh 1951 a)

that the main effect of alloying is to change the number of occupied states in the band scheme. These calculations were carried out by Leigh, whose results may be summarized in the form of fig. 39. The full curves represent the effect of variation of the contributions from the overlapping electrons alone, while in the broken curves the change in the coulomb contribution has been included. The large rise in $(C_{11}-C_{12})/2$ as the electron concentration falls is due to the decrease in the magnitude of the negative term arising from electron transfer between the square faces when the appropriate shear takes place. The rise in C_{44} is due to the number of electrons overlapping across the hexagonal faces becoming small. The peak in the curve for $(C_{11}-C_{12})/2$ is reached at 2·67 electrons per atom, which is where the overlap across the square faces ceases. Further decrease in the electron/atom ratio leads to the production of holes in the first Brillouin zone which cause a sharp fall in the curve.

In connection with figs. 38 and 39, it is very significant that Raynor and Wakeman (1949) showed that the primary solid solubility boundary for silver

in aluminium possessed a sharp change in direction at approximately 2·62 electrons per atom, which is exactly the electron concentration at which the primary solid solution of zinc in aluminium splits up on cooling into two separate face-centred cubic solutions. In discussion of these effects Raynor and Wakeman showed how the phenomena could in both cases be accounted for in terms of an anomaly of an electronic origin in the free-energy curve of the solid solution, and an attempt was made to relate the anomaly to the curve of the density of electronic states derived from Matyas' work. It would now appear that the probable cause of the free-energy anomaly is the sharp rise in $(C_{11}-C_{12})/2$ to a peak in the same range of electron/atom ratio. A high elastic constant corresponds with a small vibrational entropy, so that a part of the entropy of the system is anomalous at electron/atom ratios in the region of 2·67; this may well be sufficient to cause a small bump on the free-energy curve, leading to the splitting up of the aluminium–zinc solid solution into two phases.

This problem has been examined by Leigh (1951 b), on the basis of the contributions to the free energy of the electronic energy, the mixing entropy and thermal vibrations. He points out that the first derivative of the Fermi energy with respect to the electron/atom ratio cannot have any discontinuity, on the assumption that no appreciable change in the density of states curve for aluminium is caused by the introduction of zinc. The second derivative $(1/N(E_m)$ where $N(E_m)$ is the density of states at the Fermi surface) cannot be negative anywhere, and the conclusion is drawn that the Fermi energy cannot give rise to any subsidiary maximum in the free energy per atom. The whole electronic energy varies smoothly and approximately linearly with composition and can be neglected for the present purpose. The contribution to the free energy per atom from the mixing entropy is, of course, $kT[x\ln x+(1-x)\ln(1-x)]$, while the contribution from thermal vibrations may be written

$$\int_0^\infty [\tfrac{1}{2}h\nu+kT\log\{1-\exp(-h\nu/kT)\}]f(\nu)\,d\nu \qquad \ldots\ldots(65)$$

where $f(\nu)\,d\nu$ is the number per atom of normal modes of vibration of the lattice having frequencies between ν and $d\nu$. Since only the effect of changes in elastic constants on the frequency spectrum is required, and not the actual form of the spectrum, it is legitimate to neglect the effect on the spectrum of the dependence of velocity of propagation on wavelength, which may be assumed to vary smoothly with zinc concentration. Taking the Debye spectrum for an isotropic crystal as a basis for calculation, and taking for μ, the rigidity modulus, the mean of $(C_{11}-C_{12})/2$ and C_{44}, the contribution of the thermal vibrations to the free energy per atom may be shown to be

$$\frac{9}{8}k\theta_D+9kT\left(\frac{T}{\theta_D}\right)^3\int_0^{\theta_D/T}\ln(1-e^{-\xi})\xi^2\,d\xi, \qquad \ldots\ldots(66)$$

where $\theta_D=(h/k)(3/4\pi\Omega)^{1/3}C$ is the Debye characteristic temperature, $\Omega=$the atomic volume, and $C=$the mean wave velocity, which is related to the density ρ, the bulk modulus K and the rigidity modulus μ by the expression:

$$\frac{3}{C^3}=\left(\frac{\rho}{K+4\mu/3}\right)^{3/2}+2\left(\frac{\rho}{\mu}\right)^{3/2}. \qquad \ldots\ldots(67)$$

C, and therefore θ_D, are calculable in terms of the elastic constants $(C_{11}-C_{12})/2$ and C_{44}, so that the vibrational contribution to the free energy may be worked out.

Figure 40 shows the sum of this term and the entropy of mixing term plotted against the atomic concentration of zinc; the broken curve represents the free energy curve which would arise taking into account density changes, but assuming μ to have the value for pure aluminium throughout. These curves have been calculated for a temperature of 548° K (275° C), and the calculated and observed phase boundaries are at the following compositions: calculated 20 atomic % Zn, 70 atomic % Zn; observed 16·0 atomic % Zn, 59·4 atomic % Zn. The agreement is as good as could be expected, taking into account the approximations involved in the calculation of the variation of the elastic constants with electron/atom ratio.

The present theory does not account for the disappearance of the solid solubility gap in the aluminium–zinc alloys at temperatures above 365° C, and a complete theory must take account of the effect of temperature on the elastic constants also. The theory is, however, of considerable interest in so far as it draws attention for the first time to the effect of anomalous vibrational entropies on the existence of phase boundaries.

The electronic distribution derived by Matyas for aluminium necessitated the presence of 'holes' in the first Brillouin zone. The effect of a distribution of holes was considered by Leigh, who concluded that, though an explanation

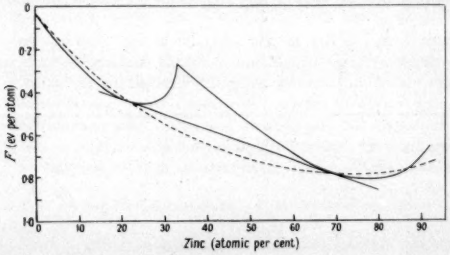


Fig. 40. Free energy plotted against composition for the face-centred cubic solid solution of zinc in aluminium. (*After* Leigh 1951 b)

of the elastic constants is possible for holes behind the {200} faces of the zone, improbably large values of electron energy at the surface of the zone need to be assumed. The elastic constants, however, cannot be explained by assuming a small number of holes at the corner points of the zone instead of an overlap across the square faces, though the presence of these holes in addition to the postulated overlap has only a very small effect on the results.

The electronic structure of aluminium has therefore been examined in fair detail, and may be said to be as satisfactorily established as the present computational methods permit.

Whether the first Brillouin zone is overlapped across both the {111} and {200} faces, or only across the {111} faces, it is clear that the overlaps are symmetrical. The effect of an electronic overlap on lattice spacings has been discussed with reference to magnesium; it is probable that a similar, but symmetrical, effect occurs with aluminium. The atomic diameter of aluminium (2·86 Å) is abnormally large compared with the diameter to be expected by comparison with neighbouring elements in the Periodic Table. This may be

attributed to the effect of the symmetrical overlap, which tends to expand the lattice of the metallic cyrstal (Jones 1934). Considerable support is given to this view by the observation (Hume-Rothery and Raynor 1940) that in many alloy systems the apparent size of the aluminium atom as deduced from lattice spacing measurements is smaller than that indicated by the interatomic distances in the crystal of the element. This phenomenon is restricted to cases in which the aluminium atom forms part of an alloy in which no Brillouin zone overlap takes place. Thus, when dissolved in copper or silver, where the number of electrons per atom in the alloy is too low for overlap to take place across the faces of the first Brillouin zone, the apparent size of the aluminium atom is approximately 2·7 Å. When dissolved in magnesium, however, where the alloy corresponds with an overlap of electrons from the first Brillouin zone in the direction perpendicular to the hexagonal axis, the apparent size of the aluminium atom as derived from measurements of the lattice spacings perpendicular to the hexagonal axis is 2·8 to 2·84 Å. It is thus comparable with the atomic size in the pure metal. It is very probable, therefore, that the abnormally large size of the aluminium atom is due to the presence of the electron overlaps discussed in this section.

§ 6. THE TRANSITIONAL METALS

On passing along the second period of the Periodic Table from sodium to chlorine, the nuclear charge and the number of electrons outside the closed shell system of neon ($1s^2$; $2s^2$, $2p^6$) increase in unit steps. Sodium, magnesium and aluminium are metallic in the crystal of the element, while silicon is an insulator (see below). Phosphorus, sulphur and chlorine are non-metals; the change in properties is due essentially to the increase in the number of electrons in the third quantum shell from unity for sodium to seven for chlorine. The addition of one more electron gives the third quantum shell a provisional stability at the distribution ($3s^2, 3p^6$) corresponding to the inert gas argon. The next period begins with potassium, which has one 4s electron outside the closed shell system of argon, and is an alkali metal. The sequence of changes in the electron distribution of the free atom, and in the properties of the element is, however, quite different on passing along this period from the simple sequence in the preceding period. Calcium has two 4s electrons outside the argon system, and behaves as a typical alkaline earth metal. The further addition of an electron to give scandium, however, takes place not into the 4p levels but into the underlying 3d levels; the 3-quantum shell, which attained a provisional stability with the configuration $3s^2, 3p^6$, becomes for scandium $3s^2, 3p^6, 3d^1$. The electronic structures of the succeeding elements are tabulated below; no element of halogen-like properties occurs until the atomic number 35 is reached, and the next inert gas structure has the configuration ($1s^2$; $2s^2$, $2p^6$; $3s^2$, $3p^6$, $3d^{10}$; $4s^2$, $4p^6$).

Titanium	$1s^2$; $2s^2$, $2p^6$;	$3s^2$, $3p^6$, $3d^2$; $4s^2$
Vanadium		$3s^2$, $3p^6$, $3d^3$; $4s^2$
Chromium		$3s^2$, $3p^6$, $3d^5$; $4s^1$
Manganese		$3s^2$, $3p^6$, $3d^5$; $4s^2$
Iron		$3s^2$, $3p^6$, $3d^6$; $4s^2$
Cobalt		$3s^2$, $3p^6$, $3d^7$; $4s^2$
Nickel		$3s^2$, $3p^6$, $3d^8$; $4s^2$
Copper		$3s^2$, $3p^6$, $3d^{10}$; $4s^1$

Thus the process of filling up the 3d levels which began at scandium is not complete until the element copper is reached; at this stage the 3-quantum shell is full with eighteen electrons. The elements from scandium to nickel are known as 'transitional metals' and are of considerable interest theoretically. Corresponding series of transitional metals occur later in the Periodic Table; these are the series from yttrium to palladium, and from lanthanum to platinum; the element actinium probably begins a fourth series which includes uranium and the transuranic elements.

The problem of the band structure of solid transitional metals cannot yet be regarded as satisfactorily solved, but a considerable amount of work has been carried out on the ferromagnetic metals of the series manganese to nickel. Of these elements, nickel has been studied in most detail, and the structures of the other metals have been discussed mainly by analogy.

The chemistry of the transitional metals is complex and they show a very variable valency in their chemical reactions. This indicates that the energies of the 3d and 4s levels in the atom are closely similar, and very little more energy is needed to ionize 3d electrons than is required to ionize 4s electrons. We may thus expect the band structure in the solid state to be complicated, with the 3d and 4s bands overlapping in energy. The general physical picture of the electronic constitution of transitional metals was provided by the work of Mott (1935) on solid nickel, which can be extended without serious modification to the corresponding elements palladium and platinum.

Working from the observation that the energies of the normal state of the nickel atom, ($3d^84s^2$), and of the lower excited states ($3d^94s^1$) and ($3d^{10}$) exist within an energy range which is less than the binding energy per atom of the metal, it is shown that, in the metal, the wave function for each atom will be representable by the sum of the atomic wave functions corresponding to several states with different electron configurations. The wave function in the atom of the solid will thus be of the form

$$A_2\psi_2 + A_1\psi_1 + A_0\psi_0 \qquad \ldots\ldots(68)$$

where ψ_2, ψ_1, and ψ_0 represent atomic wave functions corresponding to the superposition of the states ($3d^84s^2$), ($3s^94s^1$) and ($3d^{10}$) respectively. The mean number per atom of electrons in s states is then $2|A_2|^2 + |A_1|^2$. At any given moment, the metal is to be regarded as in such a condition that some of the ions have a complete 3d shell, some have one electron missing from the 3d shell, and others two missing electrons; the number of ions with two 'positive holes' is proportional to $|A_2|^2$, and the number with one positive hole to $|A_1|^2$. In nickel, magnetic evidence indicates that $|A_2|$ is much larger than $|A_1|$, and it may be considered that in general the ions have either two positive holes or none. The radius of the wave function for the d states, as in the case of copper, is considerably less than that of the s states, and the energy of interaction between the d wave functions is small. When, therefore, the solid metal is formed, atomic 3d states broaden into a much narrower band of allowed energies than the atomic 4s states. Since the atomic states do not differ greatly in energy, at the interatomic distance characteristic of the pure metal the bands will overlap considerably. This is illustrated in fig. 41 (*a*). Since the numbers of electron states to be accommodated in the 3d and 4s bands are respectively 10 and 2, it is clear that the density of states in the 3d band must be considerably greater than in the 4s

band. The curve for the distribution of states may be represented diagrammatically as in fig. 41 (*b*), and the characteristic properties of these transitional metals are due essentially to the fact that the highest energy in the Fermi distribution is less than the energy corresponding to the top of the 3d band. In the case of nickel itself, where there are ten electrons to be accommodated, the number of s electrons, and of positive holes in the 3d band, are equal, and physical evidence (especially magnetic evidence) suggests that this number is approximately 0·6 per atom. Again from magnetic evidence, and assuming that the band structure remains similar to that of nickel for the preceding transitional metals, the numbers of holes and s electrons for cobalt and iron are deduced as: cobalt, 1·7 holes in 3d band, 0·7 s electrons; iron, 2·2 holes in 3d band, 0·2 s electrons.

This interpretation of the electronic structures of the transitional metals receives support from a number of properties. In the first place, the magnetic properties are interpreted as due to the existence of holes in the d-shell, such that in the ferromagnetic condition, which is stabilized by a *positive* value of the exchange integral, a number of electrons of one spin quantum number exist without any corresponding electrons of opposite spin. According to this conception, the saturation magnetic moment, in Bohr magnetons per atom, is numerically

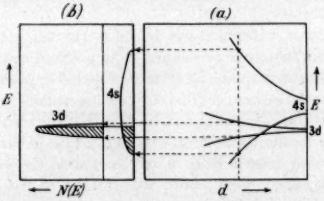


Fig. 41. Diagrammatic representation of density of states in the 3d and 4s bands of nickel. (*After* Mott 1935)

equal to the number of holes in the 3d band, so that a quantitative estimate of this number may be made. The electronic specific heats of transitional metals are also abnormally high, and this may be attributed to the presence of holes at the top of a 3d band in which the density of states is high. Thus, as noted later (page 239) the electronic atomic specific heat of a metal in which the electrons constitute a nearly degenerate gas may be written

$$C_v = \gamma T = \tfrac{1}{2}\pi^2 n_0 RT/T_0 \qquad \ldots\ldots(69)$$

where n_0 is the number of free electrons per atom, and R the gas constant. T_0 is the degeneracy temperature of the electron gas, above which the condition of degeneracy no longer holds, and which may be put equal to E_{max}/k. Since T_0 is of the order of 10^4 to 10^5, the electron gas at ordinary temperatures can be regarded as completely degenerate. A very similar formula may be applied to transitional metals with a nearly full 3d band; the only modification is that n_0 now refers to the number of holes in the d band, and the degeneracy temperature T_0 is put equal to $(E_0 - E_{max})/k$, where E_0 is the energy corresponding to the top of the band. This circumstance is due to the fact that, near the bottom of a band, the density of states is proportional to $E^{1/2}$, while, near to the top of a band, the

density is proportional to $(E_0-E)^{1/2}$. The specific heat contribution of the electrons of a transitional metal may thus be calculated for very low temperatures by the above expression, but, since T_0 now becomes of the order of 10^3, the 'gas' of 'positive holes' becomes non-degenerate at comparatively low temperatures. Although at low values of T/T_0, the heat capacity per electron is low, at ordinary temperatures it is abnormally high; since T_0 is low, the gas of positive holes may become non-degenerate at ordinary temperatures and each positive hole then contributes an amount $3k/2$ to the specific heat. At $T/T_0=\frac{1}{2}$, the specific heat per hole is approximately $1{\cdot}25k$, and the electronic specific heat is of the order of $1{\cdot}25\times 0{\cdot}6\times R$ or 1·5 cal/g atom.

Further evidence in favour of the Mott theory may be obtained from a study of the change in the saturation magnetization and Curie temperature of nickel when elements of various valencies are dissolved in it. Some of this evidence is mentioned in a later section (§ 8), but for the present purpose it will be sufficient to say that the model of the electronic structure of nickel, as proposed by Mott, serves to interpret satisfactorily most of the physical properties of the metal and of some of its alloys. On the basis of saturation magnetic moments, it would appear that similar models are appropriate for cobalt and iron, but there is little evidence to suggest that the electronic structures of manganese and chromium may be similarly interpreted.

In an earlier section, reference was made to the calculations carried out by Krutter for the band structure of copper. As pointed out by Slater (1936), it is reasonable to assume that the band system of nickel is essentially similar to that of copper, except that the system is filled with electrons only to the level denoted by L_{Ni} (fig. 19). If this is correct, then it may be seen that the results of the calculations are in good general agreement with Mott's model. The 3d band, though consisting of two peaks in electron distribution, is narrower than the 4s band, and contains approximately 0·6 positive holes above the surface of the Fermi distribution. The $K\beta_{2,5}$ emission line and the absorption edge for nickel have been examined by Beeman and Friedmann (1939) for comparison with those for copper and for comparison with the theoretical calculations. It is found that the results are very similar to those for copper in the region of high absorption. Owing to the fact that the electron levels are filled to a smaller extent, and to the non-existence of any 4s levels above the energy to which the 3d band is filled, it would be expected that the structure on the high-frequency side of the copper $K\beta_5$ line (fig. 20) would be absent for the corresponding nickel line; it will be remembered that this structure is due to emission from the filled 4s levels above the top of the 3d band. This is experimentally observed (fig. 42). The absorption edge is also similar to that for copper, but the absorption minimum (M in fig. 20) does not occur; thus the assumption that the band systems of copper and nickel are identical is not strictly true. Probably, owing to the smaller nuclear charge and the smaller interatomic distance, the 4s and 4p bands overlap more in nickel than in copper. If the absorption maxima denoted by A in figs. 20 and 42 are identified with the same peak in the density of states curve, the difference in the extents to which the band systems are filled in the two elements may be derived from the experimental results. This may be expressed in terms of the difference between the voltage separations of the points A and L in the two absorption curves, which is 2·6 volts. This value compares very favourably with the theoretical estimate of 2·4 volts (fig. 19).

As far as the x-ray emission and absorption experiments are concerned, therefore, theory and practice appear to be in reasonable agreement. Calculations have, however, recently been carried out by Fletcher and Wohlfarth (1951), who have determined the density of states for the 3d electrons in a face-centred cubic lattice, using the approximation of tight binding of electrons, with special reference to nickel. In this method, five atomic wave functions, corresponding to the fivefold degenerate 3d level of the isolated atom, are used, and the calculation is carried out in a manner somewhat similar to that described for Matyas' calculations for aluminium. The resulting provisional $N(E)$ curve, which is the sum of five constituent curves, is shown in fig. 43, which may be compared with fig. 19. It will be seen that the total width of the 3d band is only 2·7 ev though it retains the split character previously deduced. This value is only about half that suggested by Slater, and it now becomes important to decide which computational approach gives the more trustworthy results. According to the authors, a more complete investigation of the approximations involved is in progress, but it seems improbable that major differences from fig. 43 will result as a consequence of later corrections.

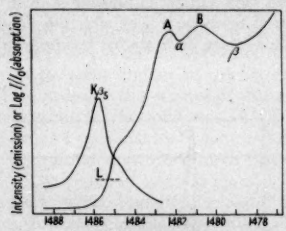


Fig. 42. X-ray absorption and emission curves for nickel.
(*After* Beeman and Friedmann 1939)

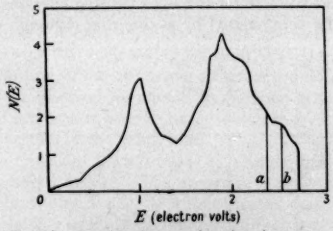


Fig. 43. Density of states plotted against energy for the 3d band of nickel.
(*After* Fletcher and Wohlfarth 1951)

Though revision of details may be necessary, the general nature of the band system for the transitional metals is now clear, and one of the outstanding features is the high density of states in the 3d band. This band is in the process of being filled up through iron, cobalt and nickel, and is full at copper. It is expected that addition of electrons to the 3d band will not raise the top of the filled levels nearly as much as the addition of the same number of electrons to the s or p bands. This has been confirmed experimentally. A measure of the alteration of the top of the Fermi distribution on adding one electron is provided, for the elements iron to germanium, by the voltage separation of the absorption maximum A and the absorption edge L for the various absorption curves (see figs. 20 and 42). The actual separations are: Fe 19·8 ev, Co 20·5 ev, Ni 17·3 ev, Cu 14·7 ev, Zn 10·1 ev, Ge 4·4 ev, Ge 4·3 ev.

From iron to nickel, the levels are within 3 ev of each other, confirming the high density of states. The figure for cobalt does not fit smoothly with the others, but this may be due to the crystal structure difference. At copper, the 3d band has filled and some electrons have energies above the top of this band, and occupy the 4s band. The energy is raised by 2·6 ev. Proceeding to zinc and gallium, the successive additions of one electron per atom raise the top of

the Fermi distribution by 4·6 and 5·7 ev respectively. These figures are entirely consistent with the general picture developed.

According to this model, the interaction between the d wave functions, which are of relatively small radius, is small, and cohesion is due essentially to the electrons with s wave functions, in much the same way as for copper. The binding energies of the transitional metals are, however, higher than the corresponding energies for copper and silver, and some additional effect must be present. As noted previously, the mean number of electrons per atom in s states is

$$2|A_2|^2+|A_1|^2. \qquad \ldots\ldots(70)$$

For nickel, this is equal to 0·6, with $A_2 \gg A_1$. Thus, at any given moment, there is 0·6 electron per atom in s-states, and, since A_1 is small, nickel ions will have either the $3d^84s^2$ configuration, with two positive holes in the d shell, or the $3d^{10}$ configuration, with a complete d shell. Two holes can exist in the same atom provided they correspond with parallel spins. While the nickel ion is in the $3d^8$ state, it will be an effectively positive ion; any positive ion can strongly attract the conduction electrons in its neighbourhood, and this additional binding is likely to account for the high binding energies referred to above. The situation may be summarized by saying that although on the average each atomic polyhedron

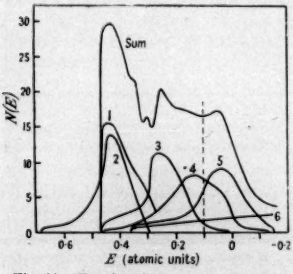


Fig. 44. Density of states plotted against energy for α-iron.
(*After* Manning 1943)

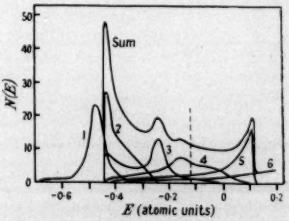


Fig. 45. Density of states plotted against energy for γ-iron.
(*After* Greene and Manning 1943)

corresponding to a nickel atom on the lattice is neutral, at a given instant a given polyhedron may contain two positive holes; this polyhedron is then positive, and the surrounding ones negative.

The type of calculation carried out by Krutter has more recently been applied to the two modifications of iron (body-centred and face-centred cubic respectively). Though these confirm the general picture of the transitional metals outlined above, they do show that the details of the band system derived for copper cannot be applied without modification to the neighbouring transitional metals. The problem of body-centred cubic iron was investigated by Manning (1943) and that of the face-centred cubic modification by Greene and Manning (1943). The details of the calculations need not at this stage be discussed, but the curves for the distribution of energy states which were deduced are of interest. These are shown in figs. 44 and 45. It will be noted that there is some similarity to the copper curve, but the details are somewhat different. Again the total 3d band is made up

of five sub-bands, and in both cases shows a certain degree of splitting; it is important to note that this band contains contributions from 4s and 4p states, as in the previous cases. The sixth, low-lying, band contains mainly 4s states, and thus confirms the general postulates of the Mott model. The extents to which the band system is filled for various numbers of electrons per atom are indicated in the figures, and it will be seen that the first two sub-bands are completely filled. The other three sub-bands of the 3d band, however, are incompletely filled, and, as in the case of nickel, it is these incompletely occupied levels which are responsible for the magnetic properties.

As a rough approximation, the distribution curves for iron may be used to obtain an idea of the band system for chromium and manganese, by calculating the area under the curves which would correspond to 6 and 7 electrons per atom respectively; it is, however, very doubtful how much reliance can be placed on this.

§7. THE BAND STRUCTURES OF OTHER METALS AND NON-METALS

One of the metals of high valency which as been investigated theoretically is tungsten, and this is of particular interest in view of the technical importance of high-melting materials. In the free atom, the electronic structure of tungsten involves an ion core of sixty electrons in closed quantum groups, outside of which are two 5s, six 5p, four 5d and two 6s electrons. In such a complex atom, it would seem probable that a considerable modification of the electronic structure would take place in the solid state. The cellular method developed by Wigner and Seitz and extended by Slater is capable of application even to heavy metals to give a reasonable approximation to the band system in the solid state.

The problem was examined by Manning and Chodorow (1939) using the general method applied by Slater to the alkali metals. The structure is body-centred cubic, so that very similar principles are involved. It will be remembered that Slater used eight atomic wave functions and satisfied the boundary conditions at eight faces of the atomic polyhedron only. The difficulty of applying this calculation to a transitional metal like tungsten lies in the fact that, in Slater's eight functions, only three of the five d functions are included. Since tungsten has a partially filled d shell in the free atom, the d bands in the metal may be expected to be partially filled, so that if only three of the d functions are used the band picture resulting will be inaccurate. Manning and Chodorow therefore satisfy the boundary conditions at all faces of the atomic polyhedron illustrated in fig. 7; this requires the use of fourteen atomic functions in setting up the wave function for a valency electron, and those chosen are one s function, three p functions, five d functions, four f functions, and a g function. Though the higher bands in the metal may be somewhat inaccurate, the results for the d and s bands should be reliable. The potential field for a valency electron in metallic tungsten was taken to be that due to the nucleus, the inner electrons, and five 5d electrons. The solutions of the boundary equations were used to provide a second approximation to the potential field, to which further approximations made little difference. From the solutions, curves of energy against k may be constructed for certain directions of propagation (the [111], [110], and [100] directions), and the results may be generalized to enable surfaces of constant energy in k-space to be drawn. It is most instructive for our purpose to consider the density of states which may be derived for the various bands, taken one at a time. The $N(E)$ curve is shown in fig. 46, in which the individual 'sub-bands' are shown together with the resultant

summation curve. The complete d-band is composed of five 'sub-bands'. The lowest band is narrow, and the energy contours referring to this band are far from characteristic of free electrons. The next two bands are wider, but the electrons still bear little relation to free electrons. The three highest 'sub-bands' are wider still, and the energy contours tend to resemble those for free electrons more strongly. The sixth band shown in fig. 46 is the s-band, and it is important to note that this band contains much less than one electron per atom; the bottom of this band lies at a somewhat higher energy than the bottom of the d-band. The three lowest 'sub-bands', particularly the lowest one, probably contribute most of the considerable binding energy of tungsten. The formation of the solid from the free atoms may be considered in terms of a formal scheme. At large atomic separations, the atoms have the configuration $5d^46s^2$; as this distance decreases, the s functions overlap, so that the single level broadens into a band of levels, while the d levels maintain their atomic character.

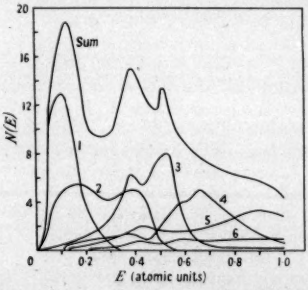


Fig. 46. Density of states plotted against energy for tungsten. (*After* Manning and Chodorow 1939)

Further decrease in the interatomic distance leads to further broadening of the s band and the d band; the s band broadens to such an extent that the upper levels have higher energies than the narrow d band, and transition of electrons from s to d band occurs. The curve representing the lowest state in the s band as a function of distance will have the form of the lowest curve in fig. 8, and it is clear that, once the minimum has been passed, the binding energy of electrons in the s band will decrease. Since the d band is only approximately half-filled, the binding energy of electrons in this band will continue to increase. Since the equilibrium distance apart of atoms in tungsten is less than that corresponding to the minimum in the curve of E_0 for the s band, it must be inferred that the gain in binding energy due to the d band over-compensates for the decrease in the binding energy of the s electrons. At still smaller distances of approach the d band will also tend to broaden in the direction of increasing energy rather than the reverse, to give rise to a repulsive force. The metal therefore has a low compressibility.

The sub-bands which make up the 3d band are very wide, and the whole d band extends over some 21·8 ev; the occupied portion corresponds with approximately 6·3 ev, so that the unoccupied portion stretches for some 15·5 ev. The

work of Coster and Bril (1943) on the fine structures of the L_{II} and L_{III} absorption edges of tungsten provide some evidence that the calculations are reliable. As the frequency in the neighbourhood of the absorption edge is increased, the absorption decreases gradually until, at the critical frequency, heavy absorption takes place with the production of a 'white line'. On the high-frequency side of the absorption edge the absorption coefficient is not a strictly monotonic function of the frequency, but fluctuates by 10% or more in a region extending to several hundreds of volts from the main edge; this is due to electrons having been excited from an inner level with sufficient energy in excess of that corresponding to the critical quantum to allow movement through the lattice, where Bragg reflection may occur. This fine structure therefore depends on the crystal structure of the metal. In the case of the L_{III} edge, the pronounced 'white line' is due to the transition of an electron from the L shell (p-states) to a hole in the d band. The width of this line corresponds to the combined widths of the unoccupied part of the d band and the width of the L level, and is estimated by Coster and Bril as 2·3 X.U. or 18 ev. Since the L level has a width of 3·4 ev (Bearden and Snyder 1941), the vacant part of the d band must correspond with a width of approximately 15 ev. This corresponds well with the theoretical band structure of tungsten suggested by Manning and Chodorow.

The results obtained for tungsten may be extrapolated with reasonable confidence to tantalum, which precedes it in the periodic table, and has one 5d electron less in the atomic state. This extrapolation may be justified by the observation, by Coster and Bril, that the L absorption edges of tantalum are very similar to those of tungsten. It may be noted that these x-ray experiments give no indication of the location of the s band. The fact, however, that the L_I edge is shifted by approximately 20 ev relative to the L_{III} edge indicates that the lowest energy state of the vacant p band is 20 ev higher than the lowest vacant state of the d band. For both tantalum and tungsten, therefore, the p band and the d band appear to be separated by some 5 volts.

Similar methods have been applied to calculations of the band structure of materials which are insulators—chiefly carbon and silicon. Since silicon has properties which are in some ways analogous to those of metals, it may be of interest to consider the theory briefly. Application of the general Wigner–Seitz–Slater method to diamond involves consideration of electrons of the L shell of the atom, for which only s and p states are occupied. A combination of eight wave functions and eight simple boundary conditions of continuity of wave function and its normal derivative at four pairs of mid-faces of the atomic polyhedron suffices for setting up the total wave function. In silicon, however, the electrons concerned lie in the M shell of the atom, and it may be expected that d states have to be considered. The geometry of the polyhedron is such that only the centres of four pairs of hexagonal faces can be used for the application of boundary conditions; thus, if more than eight wave functions are considered, additional boundary conditions become necessary. The problem has been considered by Mullaney (1944), who concluded that the best compromise was to use a combination of one s function, three p functions and three d functions for each of the two atoms in the unit cell, and, through the use of boundary conditions of continuity of the value of the function, its normal and its tangential derivatives, to obtain a solution for the xy plane. With such a complex case of fourteen functions in the combination, it was found that limitation to the propagation of

electron waves to this plane was the only feasible way of obtaining a solution. When E, the electron energy, is determined as a function of k, the wave number, it is found that to each value of k there are six values of E, and the plot of energy against distance between atoms takes the form of fig. 47. At the observed half-internuclear distance (represented by the vertical dotted line) the lowest of these values lies in the region between curves I and II. The second value lies between curves II and IV, while the third and fourth values occur in regions bounded respectively by curves III and V and curves V and VI. The regions corresponding to the fifth and sixth values are not shown in the figure as they occur at such high energies as to be of little interest. It may be noted that, owing to the approximations involved in the mathematics, additional bands of zero width arise; these in general follow one or other of the curves in fig. 47. The essential nature of the electronic structure of silicon may be understood at least qualitatively from this figure. At 0° K the eight electrons of the unit cell, considering the

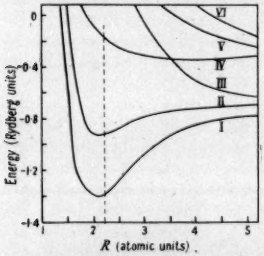


Fig. 47. Energy band system for silicon. R=half distance between nearest neighbours.
(*After* Mullaney 1944)

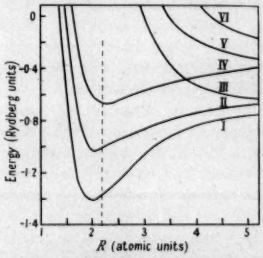


Fig. 48. Energy band system for silicon, using Kimball's approximation. R=half distance between nearest neighbours.
(*After* Mullaney 1944)

equilibrium value of the internuclear distance, occupy completely the two lower bands. The remainder are empty, and silicon is thus not expected to conduct electricity at low temperatures. Since, for R greater than 3·5 atomic units, curve IV touches the empty bands bounded by curves IV and VI, it is clear that silicon just misses being a metal; if the solid could exist at this increased lattice spacing it would presumably be a conductor of electricity.

To illustrate the effect of the additional wave functions, Mullaney recalculated the electronic structure of silicon using the combination of eight wave functions employed by Kimball (1935) in his investigation of the structure of diamond. The result is given in fig. 48, and it may be seen that there is a considerable difference between the band structures deduced, particularly with regard to the overall width of the occupied levels. Some indication of the width of these levels is given by the x-ray emission spectra examined by O'Bryan and Skinner (1934); the emission bands studied correspond with transitions from the filled levels to a relatively sharp L level, so that the width of the observed band corresponds approximately with the width of the filled levels. The observed width is $19{\cdot}2 \pm 2{\cdot}5$ volts, which compares favourably with the theoretical width of approximately 16 volts in fig. 47. The theoretical width using Kimball's

approximation (fig. 48) is approximately 9·5 volts, and thus appears to be much too small.

The theoretical calculations do not, however, agree quantitatively with conclusions drawn from experiment with regard to the width of the energy gap between the full and empty bands. From fig. 48, it may be estimated by extrapolation that the energy gap is of the order of 10 volts. The metallic appearance of silicon is due to the existence of a wide absorption band covering the range of visible frequencies; this can be ascribed to the existence of a broad, empty conduction band with a lower limit some 1·5 to 2 volts above the uppermost filled levels (Möglich and Rompe 1942) into which transitions from the filled levels are permitted. The discrepancy is thus pronounced, and it will be noted that the agreement between experiment and fig. 48 is even worse. The work of Kimball on diamond suggested a theoretical energy gap of 4 to 5 Rydberg units, which is also much too high, since the absorption spectrum suggests a gap of approximately 0·5 Rydberg unit. This disagreement is further evidence that, though the Wigner–Seitz treatment is a good approximation for the ground state of the electron and for the lower excited levels, it leads to erroneous results for states of higher energy. Thus, though the region bounded by the curves I, II

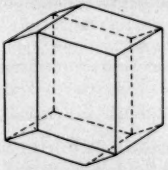


Fig. 49. First Brillouin zone for bismuth.

and IV in fig. 47 may be expected to be reasonably accurate, less reliance may be placed on the region bounded by curves III, V and VI. In reality, curve III must lie much lower at the lower distances of separation, and must be much flatter than shown in fig. 47. The average energy of the electrons in solid silicon may be calculated from the data used in constructing fig. 47, and the lowest average energy corresponds to $r=2{\cdot}35$ atomic units. The lattice spacing of silicon corresponds to 2·22 atomic units, and it is interesting to note that fig. 48, derived from Kimball's approximation, gives a lowest average energy corresponding to almost exactly this value.

It may be said, therefore, that although the theory in its present state is capable of giving an adequate qualitative picture of the more complex metals and semimetals, it is not yet quantitatively reliable for these cases.

Although the precise electronic distribution may be unknown for a polyvalent metal, valuable information with regard to the general properties of the material may often be obtained from a consideration of the appropriate first Brillouin zone. For example, the metal bismuth has a rhombohedral crystal structure, with $a=4{\cdot}736$ Å and $\alpha=57°\,16'$. The unit cell contains two atoms, at $(0,0,0)$ and $(2u,2u,2u)$ respectively, where $u=0{\cdot}237$. The zone formed by planes of the type $\{1\bar{1}0\}$ and $\{221\}$ is illustrated in fig. 49, and can contain five electrons per atom, very limited overlap from which across the $\{1\bar{1}0\}$ faces accounts adequately for many of the physical properties of bismuth. The crystal structure of bismuth

is almost that of a simple cube, which would be the case if α were 60° and u were equal to 1/4. The effect of the distortion is to introduce a structure factor for {221} planes, which is absent for the case of the simple cube. This suggests, therefore, that the reason for the distortion from simple cubic form is the tendency for the total energy to diminish as far as possible, which is favoured when the valency electrons occupy a nearly full Brillouin zone. This may be appreciated qualitatively by reference to fig. 2, which shows that the energy of electrons just below an energy gap is relatively depressed.

According to this model, bismuth, with five electrons per atom, corresponds with a first Brillouin zone which is overlapped at the points A, and has positive holes at the points B. In agreement with this, it is found that the solution of tin in bismuth, which reduces the electron/atom ratio, leads to an increase in the ratio $\chi_\perp/\chi_\parallel$ where $\chi_\perp$ and $\chi_\parallel$ are respectively the diamagnetic susceptibilities perpendicular and parallel to the principal axis. Tellurium, on the other hand, increases the overlap round the sides of the zone, and produces a decrease in $\chi_\perp/\chi_\parallel$. The magnitudes of these effects are consistent with a number of overlapping electrons of the order 10^{-4}. The model also accounts qualitatively for the electrical resistances of bismuth alloys; small additions of the four-valent metals tin and lead increase the resistance and produce an anomalous temperature coefficient so that an increase in temperature causes an increase in conductivity, while the six-valent metals selenium and tellurium decrease the resistance, and the alloys show a normal temperature coefficient of resistance.

The physical properties of certain electron compounds may also be similarly interpreted in terms of the appropriate Brillouin zone.

§ 8. THE ELECTRONIC STRUCTURE OF ALLOYS

The band theory of metals has in relatively recent years been applied to alloy formation, and perhaps the most successful of these applications is that due to Jones (1937), who has been able to interpret the experimental observation that in silver- and copper-rich alloys the primary solid solubility extends to a composition corresponding closely to an electron/atom ratio of 1·4.

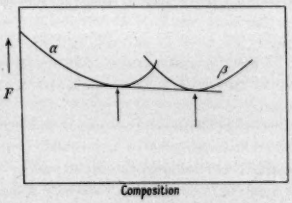


Fig. 50. Hypothetical free energy curves for the equilibrium between two alloy phases.

The phase boundaries in alloys depend upon the relative free energies of the phases which enter into equilibrium with each other. Figure 50 indicates diagrammatically the free energy curves for the α and β phases in a hypothetical alloy system at a given temperature. The points of contact of the common tangent to the two curves define the compositions of the phases in equilibrium at this temperature; an alloy with a composition lying between these two points would

have a lower free energy if it existed as a mixture of two phases than if it existed as either phase alone. The conditions for equilibrium are:

$$\left(\frac{\partial F_\alpha}{\partial x}\right)_{x_\alpha} = \left(\frac{\partial F_\beta}{\partial x}\right)_{x_\beta} \qquad \text{.....(71)}$$

and

$$\frac{\partial F_\alpha}{\partial x} = \frac{F_\alpha - F_\beta}{x_\alpha - x_\beta} \qquad \text{.....(72)}$$

where F_α and F_β are the free energies of the limiting compositions x_α and x_β of the α and β phases respectively. The slopes of the two free-energy curves are equal for solid solutions in equilibrium. Provided, therefore, there is not a very large difference between the free energies of the two phases, the α phase will become unstable with respect to the β phase at a composition in the neighbourhood of that at which the free-energy curve begins to curl upwards. For the face-centred cubic α primary solid solution of a copper or silver alloy, and the body-centred cubic β phase, such as occurs in the copper–zinc system, Jones made the reasonable assumption that all energies except the Fermi energy are almost the same for the two phases, so that the extent to which the Brillouin zones are filled in the two structures essentially controls the relative free energies at absolute zero. The curves representing the distribution of states for these two structures are shown in fig. 51 (*a*) while fig. 51 (*b*) represents the difference in energy ΔE

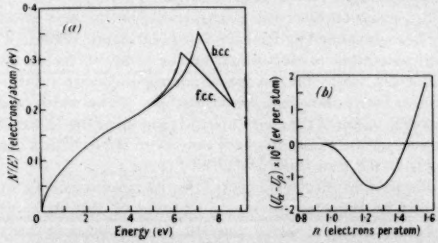


Fig. 51. (*a*) Density of states for the face-centred and body-centred cubic structures in copper-rich alloys. (*b*) Difference between Fermi energies of the face-centred and body-centred cubic structures plotted against the electron/atom ratio.
(*After* Jones 1937)

between the two structures when either is filled up to a given electron/atom ratio. It may be seen that both $N(E)$ curves are similar to the 'free electron' curve up to about 6·0 ev, which corresponds to approximately 0·95 electron per atom. The curve for the α phase then rises to a peak while that for the β phase rises more slowly. For a number of electrons per atom corresponding to this energy range, therefore, the α phase has the lower energy. With further increase of the electron/atom ratio, owing to the fall in the $N(E)$ curve for the α phase and the continued rise of the curve for the β phase, the energies tend to become equal, and at a later stage the β phase has the lower energy. These energy relations are shown quantitatively in fig. 51 (*b*). On substituting the calculated energies for the two phases in the equations defining the conditions of equilibrium, it was found that the phase boundary compositions corresponded to electron/atom values of 1·409 for the α phase and 1·447 for the β phase. The agreement of these

values with experiment provides a striking vindication of the methods adopted. Jones has also taken into account the variation of the phase boundaries with temperature by introducing entropy of mixing terms; certain aspects of the calculations involved have, however, been criticized by Zener (1947).

From this brief outline of the Jones theory, which has also been applied successfully to the interpretation of electron compounds of the γ-brass type (Jones 1934), it may be seen that it is legitimate to consider the process of alloying metals of different valency together merely in terms of the number of electrons per atom present in the resulting alloy, provided, of course, that the atomic sizes of the component metals are not too dissimilar. Thus, when zinc is dissolved in copper, each introduced zinc atom replaces a copper atom on the lattice of the solvent. Two valency electrons are therefore introduced in place of one, and the number of electrons per atom rises. More electrons are characterized by wave functions continuous through the lattice, and the Fermi energy increases. This model, which is very successful in interpreting the structures of alloys, appears to imply that the valency electrons supplied by both components are uniformly distributed throughout the metal, and it is of interest to enquire what information is available about the distribution of electrons in alloys. The x-ray emission spectrum of a solid has already been shown to give valuable information with regard to the band structure of a pure metal, so that it may be expected to throw considerable light on the case of alloys.

The x-ray $\beta_{2,5}$ emission lines and K-absorption limits for a series of copper–zinc alloys have been examined by Bearden and Friedmann (1940). The emission lines result from transitions of electrons from the 3d and 4s bands into vacancies in the atomic K levels, while the absorption corresponds with transition from the K shell into vacant levels above the Fermi surface. The width of the emission line gives directly the width of the electron band (assuming the K level to be sharp), while its shape reflects the variation of the density of states within the band, with modifications due to the transition probability factor.

From the simple free electron theory, the maximum energy in the Fermi distribution would be expected to rise on addition of zinc in accordance with the relation

$$E_{\max} = 36{\cdot}1 \left(\frac{n_0}{\Omega}\right)^{2/3} \text{ev}, \qquad \ldots\ldots (73)$$

where n_0 is the number of free electrons per atom and Ω is the atomic volume in cubic A. Assuming one and two free electrons per atom for copper and zinc respectively, $E_{\max}(\text{Cu}) = 7{\cdot}0$ ev aud $E_{\max}(\text{Zn}) = 9{\cdot}5$ ev. Since Ω changes little in the alloys, $E_{\max}$ should increase regularly with n_0, and, in the alloys, it might be expected that the emission band would widen, to the extent of approximately 1 ev at 30 atomic per cent zinc, and also that the high-frequency side of the band would move by a corresponding amount to higher frequencies. The band widths actually observed by Bearden and Friedmann were as follows:

At. % Zn in alloy	0	5	10	20	30	48	67	80	100
Widths of CuK$\beta_{2,5}$ (ev)	$8{\cdot}2 \pm 0{\cdot}2$	8·3	8·5	8·6	8·8	8·9	9·0	9·1	—
Widths of ZnKβ_2 (ev)	—	$8{\cdot}4 \pm 0{\cdot}3$	8·5	9·0	9·0	9·0	9·2	9·2	9·4

The bands for copper and zinc in a given alloy are thus of approximately the same width. For the face-centred cubic α phase, the addition of 30 atomic per cent of zinc increases the band width of copper by 0·6 ev, while that of zinc increases by the same amount. The expected increase in Fermi energy therefore appears

to occur, but not to the extent predictable from the elementary theory; further, it is observed that no marked movement of the high-frequency side of the copper emission band to higher frequencies occurs for the α brasses. A more detailed picture of conditions in alloys is therefore necessary.

The emission spectra from zinc in the alloys may first be considered. The $K\beta_{2,5}$ line is of the general form shown in fig. 25, and the $K\beta_5$ component is attributed to transitions of electrons from the 3d band. The width of this component is small and relatively constant for pure zinc and zinc in the alloys, so that the 3d levels of zinc are not appreciably broadened, and, as expected, these electrons are not shared by the structure as a whole; the 3d electrons of zinc remain tightly bound to the atom and take no part in conduction. In general, the $K\beta_{2,5}$ emission line for copper is similar to that of zinc; there is probably some 3d emission but the 3d electron band is broad enough to give no identifiable corresponding structure in the emission band. The form of the copper emission band has been mentioned earlier; it may be noted, however, that the suggestion has been made that the double 'hump' may be due to a split s band as discussed by Chodorow (1939). Although this interpretation is open to question, the emission experiments show clearly that alloying does not materially

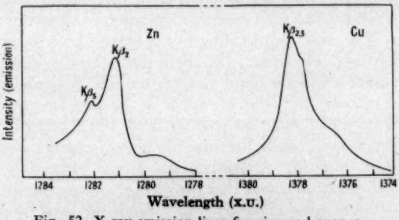


Fig. 52. X-ray emission lines for zinc and copper.
(*After* Beeman and Friedmann 1939)

affect the low energy regions of the emission bands of either component; only the high energy 4s electrons are shared by the structure as a whole, but the degree of sharing cannot be complete.

This view is confirmed by the fact that it is not possible to interpret the electrical conductivities of alloys of this class on the assumption that the electron supplied by the two sorts of atom form a uniform charge density throughout the lattice. For small amounts of zinc in copper, the electrical conductivity falls markedly. Part of the resistance is, of course, due to scattering by thermal vibrations, and is temperature dependent. In addition, the periodicity of the lattice field is further disturbed by the abnormal electronic distribution round each solute atom. If, on addition of zinc to copper, the extra 4s electron is considered to belong to the lattice rather than to the zinc atom, the increased resistance should be proportional to the scattering power of the extra field due to the zinc atoms. This problem was examined by Mott (1936), who found that, if the field round the extra charge were regarded as Ze, where $(Z+1)$ is the number of electrons outside the closed d shell for the solute metal, then the calculated resistance would be infinite. To obtain a finite resistance, a different potential $V(r)$ is required, and this potential was shown by Mott to be of the form $V(r) = +(Ze/r)e^{-qr}$,

where the quantity $q=\{4me^2(3N_0\pi)^{1/3}\hbar^2\}$, N_0 being the number of atoms per unit volume. For copper and silver, $1/q$ is 0·55 and 0·58 respectively. The charge of the zinc atom is thus screened, the electrons being drawn nearer to its position than they would be in its absence. This must mean that the greater charge of the zinc atoms is shielded by part of their valency electrons; the zinc atoms tend to retain much of their extra negative charge distribution, rather than to give it up to the lattice.

The distribution of the charges in β brass (CuZn) has also been investigated theoretically by Mott (1937); this alloy was chosen because of its simple atomic ratio and its ordered structure at low temperatures. In this alloy, it was found that each atomic cell containing a copper atom had a total charge of $-0{\cdot}075e$ while each zinc atomic polyhedron had a total charge of $+0{\cdot}075e$. In the ordered structure, each zinc atom is completely surrounded by copper atoms, so that the charge on a zinc polyhedron, $+0{\cdot}075e$, would appear to be the maximum possible for a copper–zinc alloy; the disordered alloys will correspond to an equivalent charge on a zinc polyhedron only when zinc is present in small amount, with a high probability of being completely surrounded by copper atoms.

As noted above, the x-ray emission data are consistent with the view that some electronic charge is lost from zinc to copper in the copper–zinc alloys, but the transfer does not correspond with complete donation of all electrons to the lattice. Electrons leaving zinc, and contributing to the atmosphere round a copper atom must all be high-energy electrons. If, then, the intensity of the low-frequency portion of the copper emission band and of the zinc $K\beta_2$ emission is assumed constant, the peak intensity per atom of copper on the high-frequency side increases by about 30% of its mean value on passing from 100% of copper to very small amounts of copper in zinc, while for zinc the high-frequency emission decreases by about the same amount. Consideration of the change in the integrated intensities due to 4s emission shows that the change in *integrated intensity* for zinc produced by alloying is approximately 5%. Since the total integrated intensity under the $ZnK\beta_2$ line is two electrons for pure zinc, a loss in intensity of 5% is equivalent to a loss of 0·1 electronic charge. This estimate is in good agreement with the value suggested by Mott.

One further piece of evidence that the introduced zinc atom does not give all its extra electron to the lattice is provided by the fact that the position of the peak of the $K\beta_5$ line for zinc in the alloys does not change appreciably relatively to the high-frequency side of the $K\beta_2$ component. From atomic spectra terms, it would be expected that the separation of the 3d levels (given by $ZnK\beta_5$) and 4s levels (given by $ZnK\beta_2$) would change by about 2 ev if all the zinc electrons became 'free'. If, however, the maximum loss of electrons by zinc is only approximately 0·1 electron, as suggested by Mott, the relatively constant separation experimentally observed may be understood.

It is of interest to note that the typical K absorption curve for copper (fig. 20) is also affected by the presence of zinc. The small peak M, which was interpreted as due to the small overlap between the 4s and 4p bands, gradually disappears as zinc is added, and is not present at 30 atomic per cent of zinc. The addition of zinc thus has the effect of causing a greater overlap of these two bands.

The careful interpretation of x-ray emission and absorption data for alloys indicates that alloy formation is a somewhat complex process. There is no movement of the Fermi surface comparable with that expected from the simple

'free electron' theory; addition of zinc to copper does, however, lead to an increased number of high-energy electrons around the copper atoms, and a corresponding decrease about zinc atoms. The apparent extent of this rearrangement of the high energy electrons is in agreement with a picture in which the polyvalent component of the alloy maintains a very large proportion of its valency charge.

Particular interest attaches to the electronic structure of copper–nickel alloys, in view of the interesting electronic structure of nickel, and the fact that the two metals form a complete series of solid solutions with very little lattice distortion. It might be expected (see § 6) that the band structure of nickel would be very similar to that of copper, but filled to an energy corresponding with one fewer electron per atom. Alloying with copper would then result in a raising of the Fermi surface of nickel beyond the limit of the 3d band, and this should be reflected in the x-ray emission and absorption phenomena. The work of Friedmann and Beeman (1940), however, indicates that the process is much more complex, and that direct extrapolation of the copper band system to nickel is not justified. The results may be summarized as follows:

(i) The copper absorption begins at the same energy value (within a probable error of 0·3 ev) in the alloys as in the pure metal.

(ii) Except at low nickel concentrations in copper, the nickel absorption edge is also unchanged within experimental error. For low nickel concentrations, there appears to be a shift of approximately 0·8 ev from the edge for pure nickel.

(iii) Although the emission lines show small changes in shape, there is very little shift of the lines as a whole along the energy axis.

(iv) The widths of the emission lines for copper are unchanged by alloying; the width of the nickel lines increases from 7·2 ev for pure nickel to 8·8 ev at 79 atomic per cent of copper.

There is thus no clear evidence from the absorption edges for electronic redistributions of the magnitude expected. This may be due at least partly to the fact that the two metals have approximately similar numbers of 4s electrons in the pure state, requiring very little shift of the top of the Fermi surface on alloying. Also, the potential fields of a copper and a nickel atom at neighbouring lattice points are somewhat similar; since differences in the outer shells consist mainly of the extra d electrons in copper, whose wave functions will have small radii relative to those of the 4s electrons, the extra nuclear charge of copper is well shielded at distances comparable with the interatomic distance.

The interpretation of the emission lines is somewhat complex, and may be open to question; some indication is, however, given of loss of 4s electrons from copper to nickel. According to Friedmann and Beeman, the effect of alloying is to cause the disappearance of the 3d emission in the $K\beta_{2,5}$ lines; this is interpreted in terms of emission from the admixture of p-type electrons near the top of the 3d band being eliminated as the 3d band narrows. If this loss of 3d emission is accepted, then the measured half-widths for alloys are taken at a lower intensity on the line than in the case of the pure metals. The tendency is then to find the 4s emission contribution extending over too wide an energy range. An effect of this sort would account for the apparent widening of the nickel band system. In the case of copper a similar effect would be expected; that this does not happen suggests strongly that the broadening expected is compensated for by a narrowing of the electronic band giving rise to the emission line. This could only happen by a loss of 4s electrons from copper to nickel.

That the 3d bands of copper and nickel in the alloys do contract is shown by the work of Farineau (1939), and of Farineau and Morland (1938) on the Lα emission lines of copper and nickel in their alloys. The band for the alloys narrows as the distance between atoms is decreased by alloying, but the narrowing is not very noticeable until the concentration of the emitting atom is reduced below 50 atomic per cent. When nickel is present in large amounts in copper–nickel (or nickel–aluminium) alloys, the emission line for nickel actually appears to widen slightly. Farineau has attributed this to the filling up of empty states in the nickel 3d band by the valency electrons of the other component.

Although in the copper–nickel system the absorption edges of copper and nickel show little change on alloying, it has already been observed that positive effects are exhibited by the copper–zinc system, where there are considerable changes in the number of 4s electrons on passing from one component to the other. Definite evidence on the band structure of nickel–zinc alloys might thus be expected from an examination of the x-ray emission and absorption characteristics. The system has been examined by Bearden and Beeman (1940), who measured the emission lines (Zn K$\beta_{2,5}$ and Ni Kβ_5) for two alloys containing respectively 17 and 30 per cent zinc (by weight), and also measured the absorption edges. The zinc K$\beta_{2,5}$ line is, as noted previously, a double line, the peak at the lower frequency being due to 3d emission; this peak remains unaffected by alloying. The main peak of the zinc line, however, is shifted about 0·9 ev towards lower energies with respect to that of pure zinc. The line also narrows appreciably in the alloys. Within experimental error, no movement of the peak of the nickel emission line is observed, although the width of the band increases somewhat as zinc is added. These results are consistent with the loss of higher energy electrons from zinc to the empty levels in the narrow 3d band of nickel, where the density of states is very high. This interpretation is confirmed by the shift of the absorption edge for the zinc atom to lower energies, of approximately 1·3 ev, which indicates a lowering of the top of the Fermi distribution. Although it is conceivable that the change in crystal structure from hexagonal to cubic could cause a shift in the absorption edge, the fact that no such marked shift is observed for alloys of zinc in copper, where the same crystal structure change is involved, suggests strongly that high-energy electrons are lost by zinc, and absorbed into the 3d band of nickel. The high density of states in this band is responsible for the fact that electrons can be absorbed into it without greatly affecting the absorption edge or the emission line for nickel.

The data are insufficiently critical for an accurate estimate to be made of the number of electrons per zinc atom lost to the lattice, but qualitatively it may be said that it must be greater than the 0·1 electron per atom which zinc loses when completely surrounded by copper, because the shift of the zinc absorption edge when alloyed with nickel is much greater than the shift in copper–zinc alloys. On the other hand, the number must be less than the 1·4 electrons per atom which zinc would lose were its valency electrons completely shared with the other atoms in the lattice; the simple free-electron theory would thus predict a greater lowering of the Fermi energy than is actually observed. This conclusion is supported by work on the magnetic susceptibilities of these alloys by Wheeler (1939). If the 4s electrons of zinc are absorbed into the 3d levels of nickel completely, then the paramagnetic susceptibility in the alloys should approach zero at 30 atomic per cent of zinc. Wheeler showed that this was

not strictly true, and that some paramagnetism remained. Thus it appears that some of the nickel 3d states remain unfilled even when there are sufficient valency electrons in the lattice to fill them. Further, it has been noted that nickel–zinc alloys containing x atomic per cent of zinc have higher paramagnetic susceptibilities than nickel–copper alloys having $2x$ atomic per cent of copper, even though according to the simple theory the number of electrons in the incomplete 3d band should be the same in both cases. It is probable that the stronger field of the zinc atom lowers its 4s band sufficiently with respect to that of nickel to prevent the less energetic 4s electrons from escaping from their parent atom when it is largely surrounded by nickel atoms, as in the alloys investigated. After a certain loss of electrons has taken place, it appears that it is energetically more favourable for electrons to shield the zinc atom than to go into the 3d band of nickel.

Although, therefore, there appears to be little doubt that the 3d shell of nickel, and presumably therefore of its neighbouring transitional metals, will absorb electrons supplied by the other component of the alloys the process is complex and not entirely in accordance with the simple theories.

Information with regard to the number of electrons contained in the 3d band in transitional metals and their alloys may be obtained from specific heat measurements at very low temperatures. It will readily be appreciated that at high temperatures the contribution of the electrons to the specific heat of a metal is very small; at low temperatures, however, it becomes an appreciable contribution to the whole. According to the Fermi–Dirac statistics the probability that a given quantum state of energy E is occupied is given by the expression

$$f(E) = \frac{1}{\exp\{(E-\zeta)/kT\}+1} \qquad \ldots\ldots(74)$$

where ζ is a constant which is calculable, and is approximately equal to the maximum electronic energy E_{max} at the absolute zero. We may write $\zeta_0 = E_{max}$, and it may then be shown that, at temperatures much lower than ζ_0/k the number of electrons in excited states is proportional to kT/ζ_0. The mean energy of excitation is proportional to kT, so that the energy of the 'electron gas' due to thermal motion at a temperature T will be proportional to $(kT)^2/\zeta_0$. The specific heat, per electron, is obtained by differentiating with respect to temperature, and so must be proportional to k^2T/ζ_0. A more detailed analysis shows that the total energy of the electrons (taking the lowest energy of the Fermi distribution as zero) may be expressed in the form

$$2\int_0^{\zeta_0} N(E)EdE + \tfrac{1}{3}\pi^2 N(\zeta_0)(kT)^2 \qquad \ldots\ldots(75)$$

which, on differentiation, gives

$$C_v = \tfrac{2}{3}\pi^2 k^2 T N(\zeta_0)/\rho \qquad \ldots\ldots(76)$$

where $N(\zeta_0)/\rho$ is the number of electronic states per unit energy range at the surface of the Fermi distribution for unit mass of metal; ρ is the density. Thus the electronic specific heat is proportional to the absolute temperature. The specific heat due to the thermal vibrations in the lattice is of course proportional to T^3, so that it is in general possible to express specific heat data in terms of an expression $C_v = \gamma T + bT^3$, where γ and b are constants. The coefficient γ is a measure of the electronic contribution, which, as noted above, depends upon the number of energy levels per unit energy range at the top of the band. In agreement with the fact that the density of states is much larger at the top of the filled region of the 3d

band for nickel than it is at the head of the 4s distribution of copper, the γ values for nickel and copper are respectively $1{\cdot}75 \times 10^{-3}$ and $1{\cdot}78 \times 10^{-4}$ (Keesom and Clark 1935, Keesom and Kok 1936). Since the copper–nickel alloys form a complete series of solid solutions, we might expect the shape of the nickel band system to remain relatively constant (subject to the indications of the x-ray emission and absorption experiments) and the result of adding copper would then be to fill up the 3d band of nickel as suggested by the variation of the magnetic susceptibility of nickel as copper is added. According to this interpretation, it would be expected that the coefficient γ would decrease on adding copper until the 3d band was filled at 60 atomic per cent of copper. For high copper contents, the value of γ should be similar to that for pure copper.

The system has been examined in some detail by Keesom and Kurrelmayer (1940), who determined the specific heat of four copper–nickel alloys at temperatures from 2·5° to 20° K. As expected, the specific heats followed a law of the type

$$C_v = \gamma T + bT^3. \qquad \ldots\ldots(77)$$

On plotting C_v/T against T^2, the experimental results were adequately represented by a straight line of slope b and intercept γ. From the constant b the Debye characteristic temperature may be calculated. The results of the experiments may be summarized:

Specimen Composition (At. % Ni)	100	81·61	61·97	42·07	21·58	0
γ ($\times 10^3$)	1·74	1·58	1·52	1·66	0·457	0·178
b ($\times 10^6$)	6·60	8·11	9·74	8·23	12·28	12·90
θ (°K)	413	386	363	384	335	330

The values of γ are shown plotted against composition in fig. 53; data for the nickel–iron system are included for comparison, the horizontal scale being adjusted to correspond with the same rate of change of electron/atom ratio as for the copper–nickel system. The figure is quite different from the curve which would be expected from the simple theory. Instead of decreasing uniformly to low values at 60 atomic per cent of copper, γ has a maximum at this composition, and even at 80 per cent of copper, γ has a value more than 2·5 times the value for pure copper. The constant b has a minimum at 60 atomic per cent copper, corresponding to a maximum value of θ.

The only possibility for explaining the high heat capacity of the alloys, if the simple interpretation is correct, is in terms of a specific heat of ferromagnetic origin. Keesom and Kurrelmayer show that, in their experiments, an appreciable contribution due to demagnetization must be very small. Thus, for the alloy with 42·07 atomic per cent of nickel, in order to obtain a value of θ lying on a smooth curve through the other points, it would be necessary to assume that the magnetic contribution was large at 1° K, but had become small at 20° K. This is in conflict with experiment, since more than half the estimated magnetization at 0° K was still observed at 20° K. For alloys lying at compositions above 60 atomic per cent of copper, magnetic contributions of appreciable size could not possibly be present. The abnormally high characteristic temperature for the 42·07% nickel alloy is to some extent confirmed by the elastic constants of an alloy containing 44 atomic per cent of nickel, as determined by Landon and Davies (1938), from which a value of $\theta = 387°$ K may be derived. This is in good agreement with the value of 384° K deduced from the b values of Keesom and Kurrelmayer, and it appears probable that the results do represent the true variation of γ and θ with

composition. The electronic specific heat data do not, therefore, agree completely with the predictions of the band theory. It is difficult, on the basis of these results, to understand why the Curie point and the saturation magnetization tend linearly to zero at 60 atomic per cent of copper; the specific heat experiments would suggest that the 3d band of nickel shifts relatively to the 4s band as copper is added, so that the filling of the former is not complete at the 60 atomic per cent limit. An explanation of this type might also account for the strong paramagnetism of the copper–nickel alloys in which the copper content exceeds 60 atomic per cent which is difficult to interpret according to the band theory.

In the present state of the theory of transitional metals and their alloys, little can be said of the results referring to the nickel–iron system in fig. 53. It is, however, of some interest to note that the γ values for the equiatomic iron–nickel alloy and for pure cobalt are very similar, suggesting that the band system is similar in the two cases. There is, however, a large difference in the characteristic temperatures.

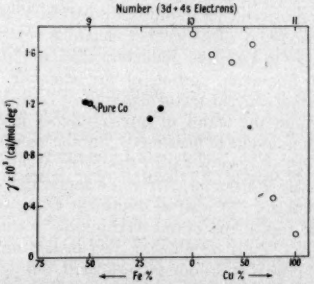


Fig. 53. γ plotted against composition and number of (3d + 4s) electrons for nickel–copper and nickel–iron alloys. (*After* Keesom and Kurrelmayer 1940)

From the experimental evidence available, and summarized in this section, it is clear that the extension of the band theory of pure metals to alloys does not accurately interpret all the effects observed. Although an adequate picture of the formation of alloys is given by the conception that, where the atomic sizes of the component metals do not differ much, the stability of alloy phases can be interpreted by assuming that the Fermi energy of the solvent is raised or lowered by the electrons belonging to the solute metal, the actual condition of the alloy must be different. The band theory, strictly applied, would imply that all the valency electrons were evenly shared between the atoms, giving a uniform charge distribution throughout the lattice, which would be composed of ions bearing their normal positive charges. The electron density round each type of atom, however, cannot be as similar as the simple theory would imply. Experimental results indicate that in nearly all the binary systems studied, the valency electrons, which must be taken as contributing to the band system, distribute themselves quite differently around the two constituents of the same alloy. It is true that, according to Farineau (1938), aluminium and magnesium in the intermetallic compounds Al_2Mg_3 and Al_3Mg_2 give rise to identical x-ray emission lines, but

this behaviour appears to be the exception rather than the rule. In general, there are considerable differences not only in the general appearances of the emission lines from the two components, and the corresponding absorption edges, but also in the energy breadth of the filled levels deduced for the two elements from the widths of their emission lines. For instance, Farineau (1939) has shown that, in the aluminium–copper and aluminium–nickel alloys, the bands for the component metals maintain approximately the same widths throughout the whole alloy series as in the pure elements. The only evidence for the loss of high-energy electrons from aluminium to copper or nickel is the reduction, for the alloys, of the intensity of the high-frequency portion of the aluminium emission line, and a slight broadening of the copper and nickel lines. At the same time, however, there is no marked shift of the high-energy limit of the aluminium emission and therefore no strong evidence of the expected lowering of the absolute energy of the top of the Fermi distribution by alloying. The behaviour is thus very similar to that described in detail for the copper–zinc alloys, and it appears very probable that, in all cases, the introduced polyvalent ion is shielded by a concentration of 'lattice' electrons, leading to a much smaller residual positive charge around the polyvalent ion than would be the case if all its valency electrons became effectively 'free'. As a consequence there is a smaller reduction than would be expected in the intensity of the high-frequency portion of the emission line. The picture indicated, therefore, is that when metals of high and low valency are alloyed together, the charge density round the atoms of higher valency is decreased somewhat compared with what it would be in pure metal, but the energy of the highest filled states for the higher valent metal is hardly decreased at all. A complete theory must take account of these indications. It would appear that the differences in the average internal potentials are at least as important as are the differences in the kinetic energies at the top of the Fermi distribution in controlling the mutual adjustment of the filled levels on alloying two metals together. Ordinarily, only the effect of the Fermi distribution is considered, but the results suggest that the lower average potential of the high valency component of a binary alloy lowers the energy of the whole band at approximately the same rate as its extra valency electrons raise the kinetic energy of the highest filled levels. According to this conception, the failure to find any marked shift of absorption edges or emission lines, and the existence of emission lines with quite different widths from the two components of the alloy may be interpreted; the electrons around the atom of high valency may be regarded as having a wider distribution in energy because of the lower potential. For alloys with transitional metals, there is the additional complication of the existence of the partially filled 3d band, which cannot be assumed to remain unaltered in energy or width as alloying proceeds. For example the empty states on the 3d band of nickel may be energetically favourable to permit considerable ionization of an element of higher valency and absorption into the empty nickel states, but this may not take place to the same extent in different alloys. Thus an actual lowering of the Fermi distribution was observed in the nickel–zinc alloys, while the experimental indications for the copper–nickel system were much less definite.

It has recently been suggested that when, for instance, aluminium is alloyed with copper, the three valency electrons of the solute metal should not be regarded as completely donated to the lattice, thereby raising the Fermi distribution as

postulated in the simple theory. Instead, only one electron should be regarded as given up to the Fermi distribution, which is thus little altered in energy or width, while the other two enter 'bound levels' beneath the Fermi distribution. The electrons in bound levels will have wave functions of a different type from the effectively 'free' electrons in the Fermi distribution. At first sight it is difficult to reconcile this picture with the success of the Jones theory of the stability of electron compounds, but the problem is under consideration, and it would appear that some mechanism for introducing 'forbidden states' into the first Brillouin zone may be necessary to avoid conflict between the various theories, which have, individually, much to commend them.

§ 9. CONCLUSION

In the confines of a review of this nature many topics of great practical and theoretical interest have necessarily to be omitted. Thus, no mention has been made of the theory of ferromagnetism, the theory of electrical conductivity, the behaviour of semiconductors, or of superconductivity, to mention but a few.* It must be admitted that certain of these topics raise special problems, but in general it may be claimed that the phenomena observed are consistent with the band theory of metals as outlined here.

Qualitatively, the nature of metallic substances is now clear; it is in the quantitative application of the theory that difficulties arise, owing to the necessity for the introduction of approximational methods, and correction factors which may in many cases have a semi-empirical nature. The nature of the agreement between theory and experiment in the case of the monovalent metals of groups IA and IB of the Periodic Table suggests strongly that the electronic nature of these metals is correctly understood. Even here, however, complications arise in the calculation of exchange and correlation energies, which are necessary as correction terms when any attempt is made to supersede the model of the electron gas in a periodic potential and to give more careful attention to the electronic interactions.

Owing to the difficulties involved in the attempts to treat the problem of correlation accurately by Wigner (1934, 1938), Landsberg (1949) has suggested that the inter-electronic terms in the energy could be simulated by replacing the coulomb potential in the exchange integral by a screened potential:

$$V(r_{12}) = e^2 r_{12}^{-1} \exp(-\lambda r_{12}). \qquad \ldots\ldots(78)$$

This empirical method has been applied to soft x-ray emission spectra by Landsberg, and its application to the specific heat of free electrons in metals has been discussed by Wohlfarth (1950); such a method is, however, still not completely satisfactory for calculating the cohesive energies of metals, because the screening term affects the coulomb as well as the exchange energy term in the total energy expression.

It is clear that the present picture is adequate for a reasonably quantitative interpretation of the compressibilities and the elastic constants of the monovalent metals, and explains in an elegant manner the marked anisotropy of the elastic properties. Accuracy is not, however, yet obtained in the consideration of such properties as the paramagnetic and diamagnetic susceptibilities.

* See Stoner (1946–7) for a discussion of the collective electron treatment of ferromagnetism; the subject of semiconductors is treated by Shockley (1951).

When divalent metals are considered, the difficulties in computation and the approximations necessary increase markedly. The detailed calculations of Herring and Hill (1940) for beryllium indicate the considerable care which must be exercised in attempting to obtain quantitative results, and especially illustrate the difficulty of making exchange and correlation corrections which refer to electrons in the state in which they occur in the metal, rather than to electrons considered to be effectively free. Similarly, the calculations for magnesium illustrate corresponding differences, and, although reasonable agreement between theory and experiment is obtained, the computations involve the assumption that the valency electrons are effectively free, and attempts to correct for this assumption tend to make the agreement worse. One of the major difficulties lies in the necessity to consider, for accurate results, the overlapping of electrons from the first Brillouin zone which occurs for the divalent metals; at present this cannot satisfactorily be done, though approximational methods are being developed which may make a better approach possible.

In calculations involving any estimate of the energies of electrons in states of higher energy, where the wave vector **k** differs considerably from zero, reliable values are still difficult to obtain quite apart from the question of overlap. Although Slater's cellular method is capable of giving reasonably accurate results when the boundary conditions are applied at many points on the surface of the atomic polyhedron, the calculations necessary are almost prohibitively heavy. It is therefore of interest to note that Von der Lage and Bethe (1947) have modified this method in order to obtain good results with a practicable amount of calculation for points of high symmetry in the Brillouin zone.

The case of magnesium illustrates the profound effect which the nature of zone overlaps may have on the alloying properties of metals, and it is now possible to account quantitatively for the lattice spacings of magnesium-rich alloys on the basis of such overlaps. The exact nature of the overlaps cannot, however, be analysed with certainty, and physically probable assumptions must be relied upon.

For aluminium, considered as an example of the three-valent metals, the mathematical difficulties are even more serious; calculation indicates that the overlap, in this case approximately one electron per atom, takes place across the {111} faces of of the first Brillouin zone, but not across the {200} faces, whereas the interpretation of the isotropic character of aluminium involves the very probable assumption of overlap across the {200} faces also. The demonstration that the details of the Fermi distribution in the second zone are responsible for the modification of the elastic constants from those which would be expected on the assumption of free electrons is of considerable importance. The variation of elastic constants with electron/atom ratio enables certain phenomena encountered in aluminium-rich alloys to be interpreted, and it may be claimed that the electronic structure of aluminium is now understood. No similar calculations have been made for gallium, indium and thallium, where complications due to complex crystal structures and to possible incomplete ionization in the solid state arise; for these metals only qualitative deductions are possible.

Knowledge of the electronic structures of the metals of higher valency, and particularly of the transitional metals, must remain somewhat qualitative, though considerable success has been obtained in interpreting their general properties, which indicates that the models adopted are reasonably appropriate. At the same

time, some details remain obscure. In this connection, it is of interest to note a recent contribution by Mott (1949) referring to the main approximations employed in dealing with metals of a transitional type. Mott points out that the collective electron treatment, which is largely used in the discussion of metals, and the London–Heitler treatment, which, in essence, starts from atomic wave functions and is difficult to apply to metals, are not merely different approximations to the same exact wave function. They represent treatments appropriate for different classes of material, and to illustrate this the case of nickel oxide is instanced. In this oxide the nickel is present as ions of configuration $3d^8$, and the 3d state should thus give rise to a band capable of containing 10 electrons, though split by the cubic field into two sub-bands. According to the collective electron picture at least one of these sub-bands should be only partially filled, thus giving rise to metallic conduction; nickel oxide is, however, an insulator, which would be consistent with the other treatment. For conduction, according to the second approximation, Ni^{3+} and corresponding Ni^{1+} ions must be produced; since this needs a relatively large acquisition of energy, no such pairs of ions having one extra electron and one fewer electron than the configuration $3d^8$ are present except possibly at very high temperatures. Where it is necessary to consider the existence of extra electrons

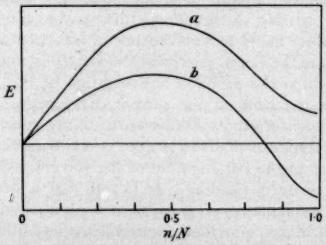


Fig. 54. Hypothetical plot of energy against the number of 'pairs' n: a, for large atomic separation, and b, for small atomic separation as in metals. N = number of atoms. (*After* Mott 1949)

coupled with positive 'holes', the London–Heitler approximation breaks down, and the collective electron treatment must be used. Since work has to be done to separate the electron and the 'hole' constituting the first pair to be formed, because they attract each other with a force derived from a potential energy $-e^2/\kappa r$ where κ is the dielectric constant, a state of the material with a small number of holes has a higher energy than one with no holes. On the other hand, if some pairs already exist, work need not necessarily be done to form more, since the field between the electrons and holes is now to some extent screened, and is of the form derived from a potential of the type $(-e^2/\kappa r)e^{-qr}$, where the constant q increases with the number of holes and electrons present. If q is large enough, electrons and holes are no longer bound to each other in stationary states, as they are in the case of a few such pairs. It is thus suggested that, if E is the energy value obtained from a wave function of the Heitler–London type, with ionized states (i.e. extra electrons on some ions and positive 'holes' associated with others), its variation with n, the number of pairs of electrons and holes, will be of the general form of fig. 54. E

first rises as n increases, and then falls. If the interatomic distance is large, E, on decreasing, probably never reaches the value of $n=0$ (curve *a*): in the case of metals, however, where the interatomic distance is relatively small, curve *b* is suggested as more appropriate. If E in fact drops below the value of $n=0$, the number of positive holes becomes large, and the approximation breaks down; the collective electron treatment is then more appropriate. For the metal nickel, the situation resembles that summarized by curve *b* in fig. 54. Experimental evidence from many sources shows that a minimum in the energy is reached when 0·6 hole per atom have been formed, giving 0·6 conduction electron. Thus nickel is described as having 0·6 conduction electron described by periodic wave functions, and 0·6 hole, which must also be described by periodic wave functions extending through the lattice; the alternative approximation is invalid. The model is essentially similar to the overlapping band model described earlier, but emphasis is now given to the applicability of the collective electron treatment for the holes, rather than for the electrons, of the 3d band. The ideas discussed by Mott have an interesting extension to the interpretation of the x-ray absorption spectra of different classes of solids, for which reference may be made to a paper by Cauchois and Mott (1949).

Although the electron theory of metals allows a reasonably quantitative prediction of their properties, the situation with regard to alloys is less satisfactory. The addition of a solute metal to a solvent metal has been considered mainly in terms of an alteration to the Fermi energy of the electrons, and many other important factors, among them the probable modification of the band system of the solvent metal by the inclusion of the solvent, have been, if not ignored, very incompletely taken into account. As shown by x-ray emission and absorption experiments the actual state of affairs in an alloy is very complex, and the consideration of alloys in terms of the band system of the solvent is incomplete, and may even in certain cases be misleading. It is felt that a new approach to this problem is required, and it may prove possible to apply a calculation of the Wigner–Seitz–Slater type to alloys, provided the mathematical difficulties can be surmounted. Some approaches to this type of work, for alloys of the monovalent noble metals with each other, have already been attempted, but much more work is necessary before alloys of polyvalent metals with each other can be considered.

It will have been noted that throughout this review the discussion of the electronic structures of metals and alloys has been in terms of an assumed crystal structure. The theory has not yet proved quantitatively powerful enough to predict the precise crystal structure which a given metal will assume. The free energy differences which cause transformation from one structure to another are known to be small, and it would appear that, unless a considerable increase in accuracy is achieved, such small differences are likely to remain within the limits of accuracy of the methods.

REFERENCES

BARDEEN, J., 1938, *J. Chem. Phys.*, **6**, 367.
BARNETT, S. J., 1931, *Phil. Mag.*, **12**, 349.
BEARDEN, J. A., and BEEMAN, W. W., 1940, *Phys. Rev.*, **58**, 396.
BEARDEN, J. A., and FRIEDMANN, H., 1940, *Phys. Rev.*, **58**, 387.
BEARDEN, J. A., and SNYDER, T. M., 1941, *Phys. Rev.*, **59**, 162.
BEEMAN, W. W., and FRIEDMANN, H., 1939, *Phys. Rev.*, **56**, 392.
BITTER, F., 1930, *Phys. Rev.*, **36**, 978.

BLOCH, F., 1928, **52**, 555.
BRILLOUIN, L., 1931, *Quantenstatistik* (Berlin : Springer).
CAUCHOIS, Y. and MOTT, N. F., 1949, *Phil. Mag.*, **40**, 1260.
CHODOROW, M., 1939, *Phys. Rev.*, **55**, 675.
COSTER, D. and BRIL, A., 1943, *Physica*, **10**, 391.
FARINEAU, J., 1938, *Ann. Phys., Paris*, **10**, 20; 1939, *J. Phys. Rad.*, **10**, 327.
FARINEAU, J., and MORAND, M., 1938, *J. Phys. Rad.*, **9**, 447.
FLETCHER, G. C., and WOHLFARTH, E. P., 1951, *Phil. Mag.*, **42**, 106.
FRIEDMANN, H., and BEEMAN, W. W., 1940, *Phys. Rev.*, **58**, 400.
FUCHS, R., 1935, *Proc. Roy. Soc.* A, **151**, 585 ; 1936, *Ibid.*, **153**, 622.
GREENE, J. B., and MANNING, M. F., 1943, *Phys. Rev.*, **63**, 203.
GOOS, F., 1936, *Z. Phys.*, **100**, 95.
GORIN, E., 1936, *Phys. Z. Sowjet*, **9**, 328.
HARTREE, D. R., 1928, *Proc. Camb. Phil. Soc.*, **24**, 89.
HARTREE, D. R., and HARTREE, W., 1935, *Proc. Roy. Soc.* A, **149**, 210 and **150**, 9.
HERRING, C., 1940, *Phys. Rev.*, **57**, 1169.
HERRING, C., and HILL, A. G., 1940, *Phys. Rev.*, **58**, 132.
HUME-ROTHERY, W., and RAYNOR, G. V., 1938, *J. Inst. Metals*, **63**, 227 ; 1940, *Proc. Roy. Soc.* A, **117**, 27.
JONES, H., 1934, *Proc. Roy. Soc.* A, **144**, 225 ; 1937, *Proc. Phys. Soc.*, **49**, 243; 1949, *Physica*, **15**, 13 ; 1950, *Phil. Mag.*, **41**, 663.
KEESOM, W. H., and CLARK, C. W., 1935, *Physica*, **2**, 513.
KEESOM, W. H., and KOK, J. A., 1936, *Physica*, **3**, 1035.
KEESOM, W. H., and KURRELMAYER, B., 1940, *Physica*, **7**, 1003.
KETTERING, C. F., and SCOTT, G. G., 1944, *Phys. Rev.*, **66**, 257.
KIMBALL, G. E., 1935, *J. Chem. Phys.*, **3**, 560.
KOK, J. A., and KEESOM, W. H., 1937, *Physica*, **4**, 835.
KRUTTER, H., 1935, *Phys. Rev.*, **48**, 664.
LANDAU, L., 1930, *Z. Phys.*, **64**, 629.
LANDON, D. H., and DAVIES, R. M., 1938, *Phil. Mag.*, **26**, 816.
LANE, C., 1929, *Phil., Mag.*, **8**, 354.
LANDSBERG, P. T., 1949, *Proc. Phys. Soc.* A, **62**, 806.
LEIGH, R. S., 1951 a, *Phil. Mag.*, **42**, 139 ; 1951 b, *Ibid.*, **42**, 876.
LORENTZ, H. A., 1916, *Theory of Electrons* (Leipzig : Teubner).
MCLENNAN, J. C., RUEDY, R., and COHEN, E., 1927, *Proc. Roy. Soc.* A, **116**, 468.
MANNING, M. F., 1943, *Phys. Rev.*, **63**, 190.
MANNING, M. F., and CHODOROW, M., 1939, *Phys. Rev.*, **56**, 787.
MANNING, M. F., and KRUTTER, H. M., 1937, *Phys. Rev.*, **51**, 761.
MATYAS, Z., 1948, *Phil. Mag.*, **39**, 429.
MILLMAN, J., 1935, *Phys. Rev.*, **47**, 286.
MÖGLICH, F., and ROMPE, R., 1942, *Z. Phys.*, **119**, 472.
MOTT, N. F., 1935, *Proc. Phys. Soc.*, **47**, 571. ; 1936, *Proc. Camb. Phil. Soc.*, **32**, 281 ; 1937, *Proc. Phys. Soc.*, **49**, 258; 1949, *Proc. Phys. Soc.* A, **62**, 416.
MOTT, N. F., and JONES, H., 1936, *The Theory of the Properties of Metals and Alloys* (Oxford : University Press).
MULLANEY, J. F., 1944, *Phys. Rev.*, **66**, 326.
O'BRYAN, H. M., and SKINNER, H. W. B., 1934, *Phys. Rev.*, **45**, 370.
PEIERLS, R., 1933, *Z. Phys.*, **80**, 763 ; **81**, 186.
RAIMES, S., 1950, *Phil. Mag.*, **41**, 568.
RAO, S. R., and SRIRIMAN, S., 1937, *Proc. Indian Acad. Sci.* A, **5**, 343.
RAYNOR, G. V., 1940, *Proc. Roy. Soc.* A, **174**, 457 ; 1942, *Ibid.*, **180**, 107.
RAYNOR, G. V., and WAKEMAN, D. W., 1949, *Phil. Mag.*, **40**, 404.
RICHTMYER, F. K., BARNES, S. W., and RAMBERG, E., 1934, *Phys. Rev.*, **46**, 843.
RUDBERG, E., and SLATER, J. C., 1936, *Phys. Rev.*, **50**, 150.
SAMPSON, J. B., and SEITZ, F., 1940, *Phys. Rev.*, **58**, 633.
DE SAUVENIER, H., 1939, *C. R. Acad. Sci., Paris*, **208**, 1724.
SEITZ, F., 1940, *The Modern Theory of Solids* (New York : Wiley).
SHOCKLEY, W., 1951, *Electrons and Holes in Semiconductors* (London : Macmillan).

Simons, C. F. E., 1943, *Physica*, **10**, 141.
Skinner, H. W. B., 1940, *Phil. Trans. Roy. Soc.* A, **239**, 95.
Slater, J. C., 1934, *Phys. Rev.*, **45**, 794 ; 1936, *Ibid.*, **49**, 537.
Sommerfeld, A., Houston, W. V., and Eckart, C., 1928, *Z. Phys.*, **47**, 1.
Stoner, E. C., 1946–7, *Rep. Prog. Phys.*, **11**, 43 (London : Physical Society).
Sucksmith, W., 1926, *Phil. Mag.*, **2**, 21.
Tolman, R. C., Karrer, S., and Guernsey, E. W., 1923, *Phys. Rev.*, **21**, 525.
Tolman, R. C., and Mott-Smith, L. M., 1926, *Phys. Rev.*, **28**, 794.
Tolman, R. C., and Stewart, T. D., 1916, *Phys. Rev.*, **8**, 97 ; 1917, *Ibid.*, **9**, 64.
Voigt, W., 1928, *Lehrbuch der Kristallphysik* (Leipzig : Teubner).
Von der Lage, F. C., and Bethe, H. A., 1947, *Phys. Rev.*, **71**, 612.
Wheeler, M. A., 1939, *Phys. Rev.*, **56**, 1137.
Wigner, E., 1934, *Phys. Rev.*, **46**, 1002 ; 1938, *Trans. Faraday Soc.*, **34**, 678.
Wigner, E., and Seitz, F., 1933, *Phys. Rev.*, **43**, 804 ; 1934, *Phys. Rev.*, **46**, 509.
Wilson, A. H., 1936, *The Theory of Metals* (Cambridge : University Press).
Wohlfarth, E. P., 1950, *Phil. Mag.*, **41**, 534.
Zener, C., 1947, *Phys. Rev.*, **71**, 846.

GALVANOMAGNETIC EFFECTS IN CONDUCTORS

By D. K. C. MacDONALD * and K. SARGINSON †

* Clarendon Laboratory, Oxford, now at Physics Division, National Research Council, Ottawa, Canada † Somerville College, Oxford

CONTENTS

	PAGE
§ 1. Introduction	249
§ 2. The generation of the Hall field	250
§ 3. Magneto-resistance	252
§ 4. The influence of temperature on magneto-resistance	254
§ 5. The source of variation in τ and m	254
§ 6. Semiconductors	254
§ 7. The 'Justi–Kohler diagram'	255
§ 8. Rotational patterns	258
§ 9. The longitudinal magneto-resistance	258
§ 10. Size effects	259
(i) The external (or normal) size effect	259
(ii) The internal size effect	260
§ 11. The ferromagnetic metals	260
(i) The Hall field	261
(ii) Magneto-resistance	261
§ 12. Analysis of magneto-resistance	262
§ 13. Solution of the Boltzmann equation and determination of resistance	266
§ 14. Size effect	272
References	273

ABSTRACT. This review is concerned with the galvanomagnetic effects (*Hall field* and *magneto-resistance*) in conductors. The primary emphasis is on metals (including a short section on ferromagnetics), but reference is also made to semiconductors. For convenience the article is broadly divided, first, into a physical survey and, in the second part, a more mathematical treatment of some of the major problems involved; on account of the breadth of the field no claim to completeness is made.

Note.—The writers decided that it would probably be of most value if this article were written essentially in two parts: first (§§ 1 to 11 inclusive) a primarily physical survey of the field, followed by (§§ 12, 13 and 14) a more mathematical review of the problems involved.

It is hoped, however, that the inevitable slight duplication of matter involved will form no obstacle to reading the article as a whole.

§ 1. INTRODUCTION

NOWADAYS it has become almost superfluous to mention that a review of some field in physics must not be regarded as exhaustive; the authors wish to emphasize forthwith that the present article is no exception. The 'first-order' interaction between a magnetic field and an electric current, namely the HALL EFFECT, has a history of over seventy years (Hall 1879, 1880, 1881), while indeed Hall's first attempt to observe a magneto-electric interaction was to try to detect (unsuccessfully) a *variation of electrical resistance* in a german-silver wire, and in fact MAGNETO-RESISTANCE (the 'second-order' interaction)

was first observed (in the ferromagnetic metals, Fe and Ni) almost a century ago (Kelvin 1856), while Tomlinson (1883) first observed magneto-resistance in bismuth.

If concomitantly with the application of a magnetic and electric field we permit a temperature gradient also to be present then additional (thermomagnetic) effects present themselves. Thus, for example, if a transverse magnetic field be applied to a metal with a longitudinal thermal gradient then a transverse electric field is generated (Nernst–Ettinghausen effect (Nernst and von Ettinghausen 1886)). In this article we shall not discuss these phenomena, although strictly in a theoretical treatment one ought to include the possible presence of thermal gradients. However, in metals, as a consequence of the essential thermal degeneracy of the electron-gas, thermal transport effects vanish in the first approximation and one is therefore justified in treating galvanomagnetic effects as a separate topic.

No previous review of this field has been published in these Reports but an excellent survey by Sommerfeld and Frank (1931) appeared early in the history of the quantum theory of metals in the *Reviews of Modern Physics* and also Sommerfeld and Bethe (1933) included a valuable section on the subject in their article in *Handbuch der Physik* entitled "Elektronentheorie der Metalle". Apart from these, briefer surveys will of course be found in standard textbooks on the theory of metals, such as Wilson (1936), Mott and Jones (1936), Seitz (1940) and a short summary of more recent work appears in a post-war book by Justi (1948).

Apart from work on the ferromagnetics, detailed study of magneto-resistance dates essentially from the work of Kapitza (1927, 1928, 1929). It is readily appreciated that the magnitude of the resistance-change in a given metal may be expected to depend on the value of $eH\tau/mc$, that is on the ratio of the relaxation time τ due to lattice-collisions to the time of orbital rotation mc/eH in the magnetic field. Consequently, as first pointed out by Peierls (1931) and Bethe (1931) since the resistance R is inversely proportional to τ, we may expect the effect to increase as R is decreased, and of course as H is increased. With metals at room temperature the factor $eH\tau/mc$ is only of the order of 10^{-2} for $H \sim 10^4$ gauss, and consequently except in such an 'anomalous' metal as bismuth (see, for example, Kapitza 1928) the effect is exceedingly small and it is certainly not at all surprising that Hall could not himself detect it. Kapitza (1927), however, developed most ingeniously the technique of generating very high magnetic fields (up to $\sim 3 \times 10^5$ gauss) by momentarily short-circuiting a high-powered generator and was then able to survey experimentally many metals and show that a *magneto-resistive increase* occurred to a greater or less degree in all—with the exception of the ferromagnetics whose resistive variation must be discussed separately.

§ 2. THE GENERATION OF THE HALL FIELD

When a magnetic field H is applied to a simple, homogeneous, isotropic metal then initially we may assume that a force $(e\mathbf{H} \times \mathbf{v})/c$ (the 'Lorentz' force) acts on every electron. Hall himself actually undertook his research in order to show that this force did act specifically on the charge-carrier, doubting (correctly) Maxwell's statement that "....the mechanical force which urges a conductor, carrying a current, across the lines of magnetic force, acts, *not* on the electric current, but on the conductor which carries it....".

When *no* current is flowing in the conductor then evidently the average displacement of the electron gas will be zero. If, however, a current be flowing perpendicular to **H**, due to an electric field F, then there is a resultant average drift-velocity of the electron-gas, v_d, given by

$$v_d = \tfrac{1}{2}e\tau F/m \qquad \ldots\ldots(1)$$

and if, further, $\tau \ll mc/eH$, then effectively every electron will experience a net transverse force:

$$\mathcal{F} = \frac{eHv_d}{c}\left(= \left(\frac{1}{2}\frac{eH\tau}{mc}\right)eF\right). \qquad \ldots\ldots(2)$$

On the other hand when τ is comparable with, or greater than mc/eH, the actual path of an electron between collisions will be significantly affected by the magnetic field, and we can evidently no longer regard the magnetic field simply as a 'small perturbation'. This is also clear since now the transverse perturbing force $\mathcal{F}$ becomes large compared with the applied force eF.

Returning to the first case, the only net effect of the magnetic field will be to produce an initial drift of the electron stream across the magnetic field. This in turn will cause a slight excess of electron density at one edge of the conductor with a corresponding deficiency at the other, which will give birth to a transverse electrostatic field, F_H, say, tending to oppose further drift.* Equilibrium will then clearly be achieved when $eF_H = \mathcal{F} = eHv_d/c$, and thus

$$F_H = \frac{Hv_d}{c} = \left(\frac{1}{2}\frac{eH\tau}{mc}\right)F. \qquad \ldots\ldots(3)$$

Using, further, $j = Nev_d$, where j is the current density (e.s.u.), and N is the number of electrons (or, more generally, effective charge-carriers) per unit volume,

$$F_H = Hj/Nec. \qquad \ldots\ldots(4)$$

Since the *Hall coefficient* R is defined experimentally through $V_H = RHjb$, where V_H is the transverse potential observed in a specimen of breadth b carrying a current-density j, it is obvious that in the ideal case discussed above

$$R = 1/Nec. \qquad \ldots\ldots(5)$$

In table 1 are given some experimental and calculated values (Seitz 1940) assuming that N is equal to the number of *atoms* per unit volume (i.e. one ideal, free electron per atom) and that the charge-carrier *is* the electron.

Table 1

Metal	Cu	Ag	Au	Li	Na	Be	Zn	Cd	Bi
$R \times 10^{-12}$ (V/cm abamp gauss)									
Experimental	−5·5	−8·4	−7·2	−17·0	−25·0	+24·4	+3·3	+6·0	~−1000*
Calculated	−7·4	−10·4	−10·5	−13·1	−24·4	−2·5	−4·6	−6·5	−4·1

* Varies markedly with orientation of crystal axis.

It is immediately evident that the general agreement is good for the monovalent metals confirming that these metals transport electricity through practically free electrons in the proportion of about one per atom. In particular, the almost perfect numerical agreement in the case of sodium concords with the evidence from other sources (e.g. the magneto-resistance, see below) that this element conforms very closely to the idealized free-electron metal.

* It is perhaps of some interest to recognize this action as an example of 'negative feedback', in such a sense as always to reduce to zero the transverse electron flow

On the other hand, the reversal of sign in the case of beryllium, etc.—while very puzzling on a classical model—is interpreted in the modern band-theory of metals as showing that at least the major part of the current flow arises in an almost full band. In other words, 'hole' conduction is dominant in such metals. Finally, in the case of bismuth the enormous discrepancy in magnitude between experiment and calculation implies that in this metal the effective number of charge-carriers is only a very small fraction of the number of atoms, a conclusion also in agreement with other experimental evidence.

We should point out that the *Hall coefficient* will not be expected to vary very significantly with temperature although on the more complete statistical theory a small variation does arise (cf. Sondheimer and Wilson 1947)—essentially due to the fact that it is only possible to define a relaxation time τ under certain limiting conditions.

In general, then, we may look to Hall coefficient measurements to give us information about the nature of the charge-carrier and its abundance.

§ 3. MAGNETO-RESISTANCE

If now, returning to our idealized model, *every* electron has the *same* relaxation time τ and effective mass m (and hence v_d), all electrons will suffer the *same* effective Lorentz force $\mathscr{F}$ which will be balanced exactly by the electrostatic force eF_H; consequently the electron-paths will be essentially unchanged in the new dynamic equilibrium and clearly therefore the current and conductivity will be unaltered. That is to say, there will be *no* magneto-resistance.

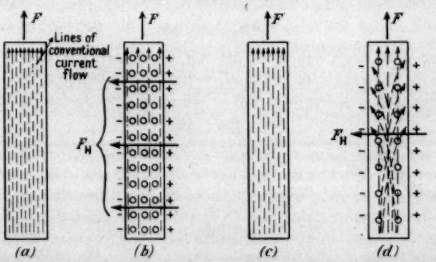


Fig. 1. Drift behaviour of current flow: (*a*) Uniform electron gas; no magnetic field; (*b*) uniform electron gas with magnetic field upwards through plane of paper, ○; (*c*) non-uniform electron gas; no magnetic field; (*d*) non-uniform electron gas with magnetic field upwards through plane of paper.

If, however, the electron-gas is not isotropic and uniform so that there is a range of values of τ (and/or m) then in the state of equilibrium after application of the magnetic field, with a steady current-flow, there will again be no *net* sideways drift of the electron stream, but because of the dependence of Lorentz force on τ, m there will be a dispersion of effective electron-paths, those with higher τ (or lower m) than average tending to drift to one side, those with lower τ to the other (see fig. 1). Since clearly both of these cases means a reduction in the contribution to the longitudinal current an *increase* in the resistance will be

observed. An elementary treatment assuming a simple distribution of two equi-probable relaxation times τ_1 and τ_2 leads then to a formula:

$$\frac{\Delta\rho}{\rho_{H=0}} = \frac{\tau_1\tau_2(\tau_1-\tau_2)^2(e^2H^2/m^2c^2)}{(\tau_1+\tau_2)^2+4\tau_1^2\tau_2^2(e^2H^2/m^2c^2)}. \qquad \ldots\ldots(6)$$

Sondheimer and Wilson (1947) have developed a rather similar formula based on an overlapping two-band model, where conduction is due partly to electrons in an 'almost-full' * (d) band and partly to electrons in an 'almost-empty' * (s) band. For small fields ($eH\tau_{av}/mc \ll 1$ in (6)) the variation predicted is quadratic in H in agreement with Kapitza's (and later) experimental results. After the initial

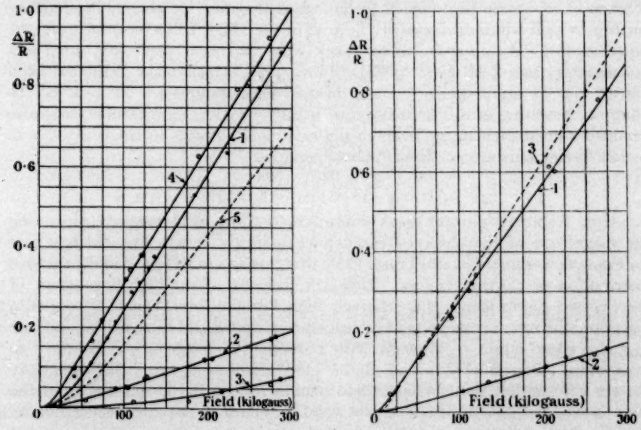


Fig. 2. (*a*) Cadmium: curve 1, H perpendicular to I, Cd_I, temperature of liquid nitrogen; curve 2, H perpendicular to I, Cd_I, temperature of solid CO_2 and ether; curve 3, H perpendicular to I, Cd_I, room temperature; curve 4, H perpendicular to I, Cd_{II}, temperature of liquid nitrogen; curve 5, H parallel to I, Cd_I, temperature of liquid nitrogen. (*b*) Gallium: curve 1, H perpendicular to I, temperature of liquid air; curve 2, H perpendicular to I, temperature of solid CO_2 and ether; curve 3, H parallel to I, temperature of liquid nitrogen. (*After* Kapitza 1929)

quadratic rise Kapitza found in general, however, a 'long' rise with H which appeared remarkably linear and in one or two cases (e.g. Te, Kapitza 1929, p. 331) a tendency towards 'saturation' was ultimately found with high fields. It is true that equation (6) (and the corresponding Sondheimer–Wilson formula) indicates (as $H\to\infty$) saturation ($\Delta\rho/\rho_0\to(\tau_1-\tau_2)^2/4\tau_1\tau_2$) and consequently also an intermediate quasi-linear region. However, it appears evident that no warrant exists for neglecting the 'direct' effect of H on the free-electron orbits as soon as $eH\tau/mc$ becomes comparable with unity, and in fact the quasi-linear region of equation (6) is generally much too limited in extent to account for the experimental behaviour (cf. fig. 2(*a*), (*b*)).

* By this we mean that the electron energies at the 'top' of each band are given respectively by $E_d = A - \alpha_1 k^2$; $E_s = \alpha_2 k^2$.

§ 4. THE INFLUENCE OF TEMPERATURE ON MAGNETO-RESISTANCE

As discussed above, the magnitude of $\Delta\rho/\rho_0$, initially at least, will be determined essentially by the value of $eH\tau/mc$. However $\rho = 1/\sigma = m/ne^2\tau$ and consequently, as first pointed out by Peierls (1931) and Bethe (1931), we may expect that $\Delta\rho/\rho_0$ will depend primarily on the ratio H/ρ_0. Since the use of very low temperatures enables a great reduction in ρ_0 to be achieved (as much as 10^{-3} to 10^{-4} below room temperature values for very pure, annealed metals), much smaller magnetic fields will then yield readily measurable effects. In fact we may expect for a reduction in ρ of 10^{-3} that fields of the order of 10^4 gauss should then produce effects equivalent to those for a field of the order of 10^7 gauss at room temperature. This must of course be qualified by the restriction that we assume dependence on H/ρ_0 which will *not* necessarily be true at very high fields (see, for example, equation (6)). As a consequence, much work in recent years (e.g. Justi and collaborators (see Justi 1948, p. 53), Milner 1937, MacDonald 1950) has been carried out at liquid hydrogen and helium temperatures ($\sim 20°\,\text{K} - 4°\,\text{K}$). In addition, however, as will be mentioned below, very low temperatures may also produce a great increase in *detail* in the rotational patterns in single crystals of metals having anisotropic electron characteristics.

§ 5. THE SOURCE OF VARIATION IN τ AND m

After Kapitza's pioneer work considerable effort went towards explaining the magnitude of the observed magneto-resistance. It was first thought (see for example Sommerfeld and Frank 1931) that this was to be found in the thermal distribution of electron energy. However, because of the high degeneracy of the electron-gas in metals (i.e. relatively high Fermi energy) this only results in a very small fractional variation of relaxation time and would furthermore indicate that the effect ought to *diminish* with reduction in temperature, contrary to experience. Peierls (1931) and Bethe (1931) first suggested the essentially correct explanation that intrinsic deformation of the Fermi surface from the ideal isotropic spherical shape would yield an effect of the correct magnitude. One would then expect, as is observed, that rotation of a specimen in a fixed transverse magnetic field would produce a varying magneto-resistance. (See, for example, Justi's book 1948 for many examples of such rotational patterns, one of which (for gold) is reproduced here as fig. 3.) Furthermore the effect of thermal agitation should now be to tend to blur over the intrinsic 'irregularities' of the Fermi surface and thus the rotational patterns should become more nearly circular with increase of temperature, as is in fact observed.

§ 6. SEMICONDUCTORS

It should however be realized that these conclusions will not in general apply to semiconductors. There the free-electron density is of course much lower than in metals and consequently the distribution is essentially classical Maxwellian. The relative thermal spread is now very considerable and will play a major part in determining the magneto-resistance which will always now be present even in an ideal isotropic semiconductor. The expression for the Hall coefficient is still of the same general form as that for metals but the numerical factor differs as a result of averaging over the thermal dispersion of electron velocities yielding now

$$R = \frac{3\pi}{8}\left(\frac{1}{Nec}\right). \qquad \ldots\ldots(7)$$

Of recent years very considerable attention has been given to this field (e.g. Crawford and Lark-Horovitz 1950, Johnson and Lark-Horovitz 1950, Klahr 1951, Jones 1951), particularly because of the rapidly increasing practical applications of semiconductors such as germanium as 'cold' diodes and transistors by Shockley, Brattain and others of the Bell Telephone Laboratories. In the papers of Lark-Horovitz, in particular, the expression for the Hall coefficient R is generalized to include two types of *scattering* in semiconductors– by the

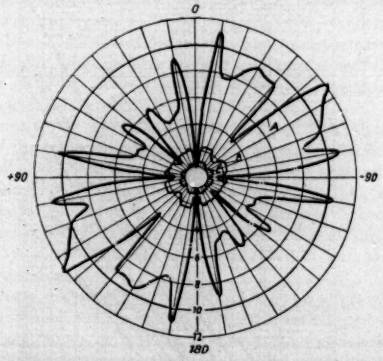


Fig. 3. Relative resistance increase $\Delta\rho/\rho$ of a gold crystal rod cut parallel to the cubic edge in a transverse magnetic field as a function of angle of rotation. Curve A, $T=4{\cdot}2°$ K, $H=18{\cdot}6$ kG; curve B, $T=20{\cdot}4°$ K, $H=35{\cdot}2$ kG; curve C, $T=20{\cdot}4°$ K, $H=21{\cdot}7$ kG.
(*After* Justi and Scheffers 1936)

lattice-vibrations and by impurity-centres—since in the former case the partial mean free path is independent of energy, while in the latter (cf. Conwell and Weisskopf 1950) it varies with the square of the electron (thermal) energy. When this spread of free paths has been taken into account it is found that the mobilities deduced for the charge-carriers are then in close agreement with those found by conductivity measurements (cf. Pearson, Haynes and Shockley 1950).

§7. THE 'JUSTI–KOHLER DIAGRAM'

Following the lead of Peierls and Bethe, Justi and Kohler (Justi and Scheffers 1938, Kohler 1938) suggested that experimental data on magneto-resistance should be plotted on a universal graph with ordinate $\Delta\rho/\rho$ and abscissa H/ρ often conveniently on logarithmic scales. It is then found (cf. fig. 4) that the metals tend fairly strongly to be grouped in classes according to the relative magnitude of the magneto-resistance. Thus the alkali metals lie lowest on the diagram, in support of our concept of a quasi-ideal metal with a nearly spherical Fermi surface for the electron-gas. On the other hand, the divalent metals, magnesium and beryllium, show an effect some two orders of magnitude greater which concords directly with the considerable variation of electron effective mass (and relaxation time) to be expected from a Fermi surface overlapping two Brillouin zones. Concomitantly, all the alkalis show very good agreement of the Hall coefficient with the simple free-electron gas formula. At the same time, however, we may note here the essential distinction between the information to be gained

from the Hall field and the magneto-resistance. The Hall field, as discussed earlier, gives us primarily an *average value* for the effective charge-carrier; thus in an ideal free-electron metal we expect agreement with equation (5), while in a metal where an almost full band plays the dominant role we find a reversal of sign in agreement with the 'hole'-conduction concept. Nonetheless, *anisotropy* of itself does not directly affect the Hall coefficient; thus Mott and Jones (1936)

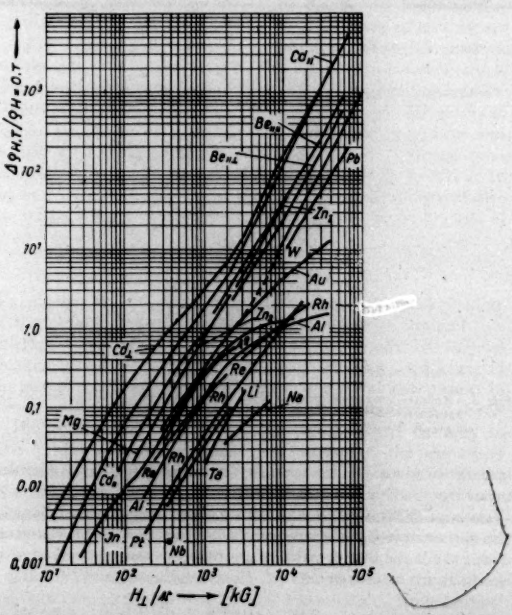


Fig. 4. Reduced Kohler diagram for different metals, comprising the results of Grüneisen, Grummach, de Haas, Justi and Kapitza. The curves for metals of the 2nd, 4th and 6th columns of the Periodic Table (e.g. Be, Cd, Pb, W) rise rapidly with $H/\bar{\tau}$ ($\bar{\tau} = \varrho_T/\varrho_{th}$) almost as a quadratic law. On the other hand the metals of the 1st, 3rd and 5th columns (e.g. Au, Al, In) tend towards a limiting value. In the 7th column Rh with an odd number of conduction electrons behaves like Au, Al, In, while Pt with an even number behaves like Be, Cd, Pb, W. (*After* Justi 1940, 1948)

point out that even a Fermi surface of ellipsoidal form, where the electron energy follows the law $E = \alpha_1 k_x^2 + \alpha_2 k_y^2 + \alpha_3 k_z^2$, will still yield the simple Hall coefficient to be expected from ideal free electrons with $E = \hbar^2 |k|^2/2m$.

On the other hand, magneto-resistance (at least for $eH\tau/mc \ll 1$) depends very essentially on the degree of anisotropy or non-uniformity of the electron-gas; it gives information, therefore, on the *fluctuation* (or deviation from the average) of electron parameters.

Returning to the Justi–Kohler diagram itself it also appears valuable for comparing the characteristics of metals belonging to the same type. Thus MacDonald (1950) has measured the magneto-resistance of all the alkali metals at low temperatures and a considerable variation is found, sodium exhibiting the smallest effect and rubidium and caesium the greatest with at least an order of magnitude difference. The alkali metals appear particularly interesting because it seems justifiable to assume, in view of their general good agreement of Hall coefficient with theory, that the conduction rests essentially in quasi-free electrons with an energy surface departing not too greatly from a sphere, the deviation being least for sodium. Jones and Zener (1934 b) obtained satisfactory agreement with Kapitza's data for lithium on the basis of an almost spherical Fermi surface; we may also note here that their theoretical values for longitudinal (see below) and transverse magneto-resistance are almost identical.

In general terms, these conclusions appear to agree with the fact that detailed measurements of the temperature variation of the electrical conductivity of the alkalis (MacDonald and Mendelssohn 1948, 1950) show that sodium agrees well with the Grüneisen–Bloch formula (cf. Wilson 1936, p. 214), while deviations become particularly great for rubidium and caesium. However, in these experiments, possible departures of the lattice vibrational spectrum from the model of the Debye continuum (Blackman 1951, Cornish and MacDonald 1951) have also to be considered.

On the other hand, Skinner (1940) concluded from his soft x-ray measurements that in both lithium and sodium the Fermi surface overlapped significantly the second Brillouin zone in strong contrast to the foregoing conclusions. It is very much to be hoped that further work, particularly on single crystals of the alkali metals, may enable a more definite and unassailable specification of the conduction electron parameters to be arrived at.

A particularly interesting application of Hall effect and magneto-resistive measurements appears to be in the field of alloys and order–disorder transitions. Two Russian workers (Komar 1941, Komar and Sidorov 1941) have investigated on these lines the $AuCu_3$ structure and the order–disorder transition around 380° c, and one of us (D. K. C. MacD.) has recently been using magneto-resistance measurements in an experimental study of some beryllium alloys (Be–Fe, Be–Ag, Be–Mg) in conjunction with Dr. Alan Blainey of the Atomic Energy Research Establishment, Harwell.

As an ordered alloy structure develops, additional symmetry will appear in the lattice (e.g. the long-range order of the 'super-lattice') which will in turn produce additional Brillouin-zone surfaces of energy discontinuity in the reciprocal (k-) space. These will in general cause distortion of the electron Fermi surface * particularly of course where it approaches the new energy surfaces, and consequently affect the Hall coefficient and the magneto-resistance. Table 2 shows some of the Russian results on the gold–copper alloy. In each experiment, after annealing at the temperature stated for the given period of time (first two columns), the alloy was quenched in water so that the state of order appropriate to the annealing temperature was then 'frozen-in'.

* A particularly striking change in a single metal occurs in α–β Sn transformation causing the white (β) tin which is an excellent metallic conductor to become a semiconductor in the gray (α) phase (cf. Busch 1951; Busch *et al.* 1951).

The rapid change of sign of the Hall coefficient around the transition temperature signifying a change from dominantly electron to 'hole' conduction is particularly remarkable, and at the same time we observe the rapid rise in magneto-resistance. The latter is much greater than can simply be accounted for by the decrease in resistivity (column 3) affecting the parameter H/ρ.

Table 2.

Annealing Temperature (°c)	Duration (hours)	Rel. Elect. Resistance	$R\times10^8$ (c.g.s.)	$\Delta\rho/\rho$ *
390	2	11·25	−645	~ 0
385	2	11·35	−631	
383	2	11·25	−637	
381	2	11·30	−640	~ 20
379	3	7·85	− 60	
377	20	7·80	+ 29	
373	50	6·73	+101	
368	21	6·35	+147	~ 46
356	24	6·05	+172	~ 80
310	67	5·13	+247	~100

* measured at −196° c for $H\simeq12$ kg.

§8. ROTATIONAL PATTERNS

As mentioned earlier, the magneto-resistance observed on a single crystal will vary to a greater or less degree as it is rotated about the axis of current flow perpendicular to a (transverse) magnetic field. Much work has been done experimentally by Justi and his collaborators (Justi 1948) and also by Dutch investigators, notably a very complete survey of gallium by Blom (1950), but the general theory of the effect is complicated and it cannot be said that the stage has yet been reached where detailed information about the Fermi surface, in particular, can be derived. The rotational patterns frequently become very complex (compare, for example, the pattern for gold at 4·2° k in fig. 3) when the effective field is strong ($eH\tau/mc\gg1$) and one might liken the general situation to the earlier days of x-ray crystallography where a wide range of possible lattice-structures had to be tried before agreement with the observed diffraction data could be achieved. The problem is further complicated here because the pattern of the Fermi surface and relaxation time is essentially capable of infinite variation as compared with the discrete nature of an atomic lattice-structure. Kohler (*vide infra*), in particular, has devoted considerable effort to the development of general theoretical principles.

§9. THE LONGITUDINAL MAGNETO-RESISTANCE

In the case that $eH\tau/mc<\sim1$ an ideal isotropic metal would exhibit no increase of resistance when a magnetic field is applied parallel to the direction of current flow since no net Lorentz force will act on the longitudinal drift-paths. If, however, the Fermi surface is distorted, the acceleration of the electrons due to the applied electric field is not necessarily parallel to the field direction; consequently the longitudinal magnetic field will act on the drift-paths of these electrons and a magneto-resistance will result, and this indeed is generally of the same order as the transverse effect. Since, as has been emphasized earlier, it is only *departure* from isotropicity which calls forth the transverse magneto-resistance (although *not* the Hall field), the latter conclusion is not surprising.

In the case of strong fields ($eH\tau/cm \gg 1$) where *direct* alteration of the electron energy levels will arise, the direction of the magnetic field relative to the applied electric field would not be expected to be very significant since of course the *intrinsic* electron-velocities (as opposed to their drift-velocities) cover all directions in space.

§ 10. SIZE EFFECTS

(i) *The External (or Normal) Size Effect*

Thus far we have tacitly assumed that our specimen of metal was 'large'—that is to say, that all its physical dimensions are great compared with the electron mean free path, l. However, when this is not so, an additional source of magneto-resistance can arise and concomitantly the Hall field—where present—is modified. On experimental grounds, low temperatures are extremely important in the field of size effects since with pure metals the electron mean free path can become relatively long (e.g. in pure sodium at $\sim 4°$ K, l may be of the order of $50\,\mu$) enabling one to deal with specimens of macroscopic dimensions. When a metal is limited in size comparable with the mean free path then in general additional scattering occurs at the geometrical boundaries and consequently the experimental specific resistivity is greater than that of the 'bulk' metal, and the relative increase becomes progressively greater as the ration d/l (d being the appropriate 'size' dimension) is reduced. J. J. Thomson (1901) was the first to treat the problem on classical lines and more recently Fuchs (1938) and others have developed the analysis on the modern statistical theory of metallic conduction.

From the physical standpoint experiments are of value in providing information about the degree of elastic scattering at a metallic surface—it is evident, for example, that were the scattering wholly elastic and specular no increase in specific resistivity would in fact arise. MacDonald (1949) found that when a *magnetic field is applied* to a sodium specimen showing size-limitation the resistivity indeed was then *reduced*, both for transverse and longitudinal directions of the field. This decrease may be interpreted crudely as an alteration of the *intrinsic* orbits of the conduction electrons by the magnetic field so that they spend less time near the surface than formerly and hence are less likely to be scattered there. The detailed analysis has been carried out by Chambers (1950), MacDonald and Sarginson (1950) and Sondheimer (1950 a, b). It follows from the foregoing that if the *intrinsic* electron velocity be v we can only expect a magnetic field to play a significant part if the orbital radius mvc/eH is comparable with the limiting dimension d; for $d \sim 5 \times 10^{-3}$ cm ($50\,\mu$), and $v \sim 10^8$ cm/sec (the estimated Fermi velocity in sodium), this gives $H \sim 10^3$ gauss. Thus again, from the physical point of view, such experiments can yield additional direct information on the intrinsic momenta of the conduction electrons.

It is found also theoretically that the Hall field (under a transverse magnetic field) is affected by size-limitation. In the case of a thin plate or film a magnetic field *in the plane* produces a Hall field whose magnitude varies across the film and is less than the value in bulk metal (MacDonald and Sarginson 1950), while if applied perpendicularly to the film-plane the Hall field—although *uniform in magnitude* (assuming no size-limitation *in* the plane) is greater than in bulk metal (Sondheimer 1950 a, b). The Hall coefficient also varies now in magnitude with the strength of the magnetic field, but is constant to a first approximation. As yet no experimental work has been published dealing directly with these Hall field predictions.

(ii) *The Internal Size Effect*

It is nowadays well known that physical metallic crystals are never perfect, suffering from 'graininess' in varying degrees. Thus, already, a considerable body of theory (e.g. see recent papers by Eshelby, Frank and Nabarro 1951, Frank 1951), is available on the subject of *dislocations* separating one region of ideal crystal structure from another. Some theoretical work has been done on the general scattering power of such crystal boundaries (e.g. Koehler 1942, 1949, Mackenzie and Sondheimer 1950), and so long as these crystal elements are small and more or less isotropically distributed they would simply contribute to a uniform reduction in the effective electron mean free path l. That is, we may write in close approximation

$$\frac{1}{l} = \frac{1}{l_{\text{th}}} + \frac{1}{l_{\text{imp}}} + \frac{1}{l_{\text{phys}}},$$

where l_{th} is the partial mean free path due to the lattice thermal scattering, l_{imp} that due to the presence of impurity atoms as scattering-centres, and l_{phys} that due to the physical imperfections of the crystal.

If, however, these crystallite regions had some preferred direction of orientation so that, for example, as a result of some previous cold-working process they tended to form a series of striations more or less parallel to the current flow we might expect to find in bulk metal effects similar to those observed when external scattering boundaries were introduced by the physical size of the specimen. In pointing this out recently MacDonald (1951) has suggested also that the *negative* magneto-resistance found in one or two metals at low temperature (particularly gold) may have its source in such an effect. It has also however been thought for some time (e.g. Meissner and Scheffers 1929; Mendoza and Thomas 1951) that small ferromagnetic impurities may be responsible for this behaviour in gold, etc., which is associated with an anomalous minimum in the resistance–temperature curve, and we should also draw attention to the very recent series of experiments by Gerritsen and Linde (1951) on silver–manganese alloys of manganese content between 0·02 and 17·8 atomic per cent. Very interesting low-temperature behaviour has been observed in this range of alloys, the resistance–temperature curve showing generally a minimum followed by a maximum at even lower temperature, while the magneto-resistive variation is strongly affected by the introduction of the manganese.

The whole problem of this anomalous low-temperature behaviour of certain metals is one of considerable interest at present and so far no adequately satisfying theory has been published.

We may also mention that deviations from Matthiessen's rule (i.e. the simple additivity of the components of resistance due to thermal and 'residual' scattering) can arise from the internal size effect.

§ 11. THE FERROMAGNETIC METALS

Although, historically, the ferromagnetics have considerable priority in this field, the mechanisms involved are still rather more obscure today than in non-magnetic metals. Three factors, at any rate, must be considered when discussing magnetic interactions in ferromagnetic metals—the influence of the 'internal' magnetic field, conveniently characterized by the intensity of magnetization I,

the spin-interaction of electrons, since fundamentally this phenomenon is responsible for the occurrence of ferromagnetism, and the occurrence of domains. In addition it may be necessary to make due allowance for the de-magnetizing factor of the specimen itself when placed in some external magnetic field.

(i) *The Hall Field*

It has long been known (cf. Kundt 1893) that the Hall field in practically all ferromagnetics, while exhibiting hysteresis and non-linear behaviour when regarded as a function of the external field H, is directly proportional—at least for low H—to the intensity of magnetization. This then has led Pugh and his collaborators (Pugh, Rostoker and Schindler 1950; Rostoker and Pugh 1951) to propose for the Hall field the relation $F_H = R_0(H + 4\pi\alpha I)$ where "... R_0 is the true Hall constant which should depend chiefly on the numbers of electrons and holes (i.e. charge-carriers) and be substantially independent of temperature.... and α is a field parameter which (is found).... to increase with temperature—from $\alpha \sim 1$ at 'low' temperatures (i.e. well below the Curie temperature) up to $\alpha \sim 100$ in the neighbourhood of the Curie point". This then implies that the conduction electrons in domains are subject to an effective Lorentz magnetic force field of the form $H + 4\pi\alpha I$. However, no detailed explanation is offered either for the large value found experimentally for α or for its variation with temperature.

(ii) *Magneto-Resistance*

When we turn to the magneto-resistance we have to take into account that electron scattering may occur in two ways—'normal' scattering by lattice vibrations and impurity centres, and also by transitions to other unoccupied electron states. Mott (1936, see also Mott 1935) in fact states that in the transition metals the probability of the essentially s-state conduction electrons making a transition to unoccupied d-states "... is several times greater than the probability of ordinary scattering...." (although this premiss has been questioned more recently), and continues "... Since the unoccupied d-states are responsible for the ferromagnetism or high paramagnetism of the transition elements, there is a direct connection between the magnetic properties and the electrical conductivity....". It is by this mechanism that Mott explains the remarkable feature that in a ferromagnetic there is generally observed a *negative* variation of resistance with magnetic field (see, for example, the most recent work of Smit 1951).

According to Mott, then, the conduction electrons will have their spins either parallel or antiparallel to an external magnetic field, and he assumes that the probability alteration during the scattering process is negligible. In the ferromagnetic state practically all the unoccupied states ('positive holes') in the d-shell are those with spin direction antiparallel to the direction of magnetization (cf. Mott 1935). It is then only possible for *half* of the conduction electrons to make transitions to d-states (i.e. those having spin antiparallel to the magnetization). On the other hand, *above the Curie point* when the spontaneous magnetization is destroyed, positive holes will have a distribution of both spin directions and *all* the conduction electrons can now make transitions to d-states. In this way Mott explains, first, the rather rapid rise of resistance with temperature on passing through the Curie point and, secondly, the decrease of resistance in a magnetic

field so long as an external field is still able to *increase* the intrinsic magnetization (i.e. at temperatures not too far below the Curie point).

We may also of course expect that a component of 'normal', positive, magneto-resistance will be present and this should *increase* with falling temperature as the scattering is reduced, while according to Mott's theory the 'abnormal' component should tend to zero as the spontaneous magnetization in zero field increases. Smit (1951) finds that he can in fact fit the experimental data at $\sim 20°$ K (on Ni and Fe) reasonably to a Kohler diagram (see § 7) if an *effective* internal field $H_{\text{eff}} \simeq H + 4\pi I$ is used in place of the external field H in the parameter H/ρ. In agreement with Pugh's conclusions for the Hall field, Smit finds that it is necessary to take the more general form $H_{\text{eff}} = H + \alpha 4\pi I$ (where $\alpha > 1$), at higher temperatures (e.g. for nickel $\alpha \simeq 2$ at liquid nitrogen temperature). Smit says also, however, that "....The explanation for the field $(\alpha - 1)4\pi I$ has not yet been given....".

We may also note that the 'abnormal' component does *not* always vanish, as expected, as $T \to 0$, and it is suggested that this may indicate that at $T = 0°$ K the spins are not completely aligned. It appears to us that the third particular feature of ferromagnetics mentioned earlier, the occurrence of domains, may have to be taken into account; one of us (MacDonald 1951) has suggested also that an 'internal size effect' may occur as a consequence of such domains and such a situation in itself could give rise (see § 10 (ii)) to an anomalous magneto-resistance.

§ 12. ANALYSIS OF MAGNETO-RESISTANCE

A correct analysis of magneto-resistance, and in fact of any phenomena in which it is necessary to determine the state of the conducting electrons in a metal, would entail solving the Schrödinger equation for the many-body system consisting of the nuclei and *all* the electrons, the presence of any external field being suitably allowed for. This is of course an extremely complicated problem and has, in fact, not been solved. As a reasonable model, however, we may take a system consisting of approximately stationary positively charged ions situated on some lattice and producing a potential field, having the same periodicity as the lattice, in which the conducting electrons move. The solution of this problem is well known, the wave function for the electrons being a plane wave modified by an amplitude which has the periodicity of the lattice. The heat motion of the ions is then considered as a perturbation effect and leads to a collision operator, which will be discussed in some detail later. Peierls (1933) has in fact pointed out that the binding forces of the lattice, assumed to be isotropic, may be allowed for by considering the conduction electrons as having an apparent mass larger than their true mass, even in the case when external fields are present. Making use of this result, Ţiţeica (1935) has found the solution for the Schrödinger equation for an electron in external electric and magnetic fields (ignoring the spin of the electrons). As in other people's work he still treats the thermal vibrations of the lattice as a perturbation. His results may be expected to be in reasonably good agreement with experiment except at extremely high temperatures or when the applied fields are very small, for this treatment requires that the disturbing effect of the lattice vibrations shall be small compared with the effects of external fields, in order that the former may, in fact, be regarded as perturbations. The analysis leads to a magneto-resistance which is proportional to H^2 for small fields and to H for large fields, in general. This compares qualitatively with Kapitza's

results, although it does not seem feasible to compare the results quantitatively at low fields for the reasons stated, namely that except at very low temperatures and in very pure metals, collisions with the lattice will have an effect comparable with, or greater than, that of the external fields, and consequently it is unreasonable to treat the former as a perturbation. Kapitza had already observed that in the presence of large fields the interactions of the electrons in the lattice and the effect of impurities were rather small, and in this circumstance we may expect Ţiţeica's results to be in fairly good agreement with experiment, provided of course we can make a reasonable estimate of some of the specific metallic constants which appear in the formulae (e.g. the apparent mass of the electrons).

In nearly all other discussions of magneto-resistance, quantization of the electron orbits in the applied magnetic field is completely neglected. This seems reasonable in the case of small fields which can be treated as perturbations, but not at all so for large fields. The validity of neglecting the quantization of orbits has in fact been discussed earlier in this article from physical principles. From the point of view of quantum mechanics we require to find the limits within which it is in fact valid to treat external fields as perturbations, this being the case so long as they produce changes in the energy and wave functions which are small compared with their unperturbed values, and further that second-order perturbation terms are small compared with first-order terms.

It is well known (Bloch 1929, see also, for example, Wilson 1936, p. 36) that in the absence of external fields and neglecting the thermal motion of the ions the admissible solution for the Schrödinger equation for the 'free' electrons in the periodic potential field of a (cubic) lattice† takes the form

$$\psi_{\mathbf{k}} = u(\mathbf{r}, \mathbf{k}) \exp\{i(\mathbf{k}.\mathbf{r})\} \qquad \ldots\ldots(8)$$

where $\mathbf{r}$ is the space vector and $\mathbf{k}$ the appropriate wave vector of the solution and the function $u(\mathbf{r}, \mathbf{k})$ has the periodicity of the lattice. In any macroscopic specimen the components of $\mathbf{k}$ may be considered as essentially continuous. The *energy spectrum*, however, consists of *bands* separated by intervals of forbidden energy, whose role in relation to the occurrence or otherwise of conductivity in specific crystals is well known. Thus the whole of $\mathbf{k}$-space is divided up into a series of zones by planes over which the energy E suffers discontinuities. These zones are called *the Brillouin zones*. In any of the zones we can reduce the vector $\mathbf{k}$ so that the products $\mathbf{k}.\mathbf{a}_l$, where $\mathbf{a}_l$ $(l = 1, 2, 3)$ are the lattice vectors, are as small as possible consistent with attaining each energy level once. In the case of a simple cubic lattice with side a we take $-\pi/a < k_1, k_2, k_3 < \pi/a$ in any zone, although of course, to obtain the *energy* it is necessary to know the zone in which the $\mathbf{k}$ vector lies.

The velocity of an electron in the state $\mathbf{k}$ is

$$\mathbf{v}_{\mathbf{k}} = \frac{h}{2\pi i} \iiint (\psi_{\mathbf{k}}^* \operatorname{grad} \psi_{\mathbf{k}} - \psi_{\mathbf{k}} \operatorname{grad} \psi_{\mathbf{k}}^*)\, d\tau \Big/ \iiint \psi_{\mathbf{k}} \psi_{\mathbf{k}}^*\, d\tau,$$

both integrals being taken over a unit cell in the crystal lattice. This can be shown quite generally to give

$$\mathbf{v}_{\mathbf{k}} = (2\pi/h) \operatorname{grad}_{\mathbf{k}} E_{\mathbf{k}} \qquad \ldots\ldots(9\,a)$$

† Sommerfeld and Bethe (1933) have considered the case of a lattice structure defined by oblique vectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$. The result is similar.

$E_{\mathbf{k}}$ being the energy in the state $\mathbf{k}$. In the case of ideally free electrons

$$E_{\mathbf{k}} = \frac{h^2}{8\pi^2 m}(k_1^2 + k_2^2 + k_3^2)$$

yielding:

$$\mathbf{v}_{\mathbf{k}} = h\mathbf{k}/2\pi m, \qquad \ldots\ldots(9\,b)$$

showing that in this case $\mathbf{k}$ is just proportional to the momentum, although in general this is not true. Nevertheless by analogy $\mathbf{k}$ is frequently referred to as the momentum vector.

The effect of small electric and magnetic fields may be determined by means of ordinary perturbation methods. We assume a (normalized) perturbed wave function $\psi(\mathbf{r}, t)$ in the form of the wave packet

$$\psi(\mathbf{r}, t) = \int\int\int g(\mathbf{k}, t)\psi_{\mathbf{k}}(x, y, z)\, d\tau_{\mathbf{k}},$$

where $g(\mathbf{k}, t)$ is to be determined, and the integration is taken over one zone only. It is important to notice that this assumes that the external fields cannot translate an electron from one energy band to a higher one. This assumption is reasonable as has been shown by Zener (1934), since in general the field required to impart sufficient energy to an electron in the time between two collisions in order to transport it across an energy gap is very many times greater than is normally used. In the first approximation magnetic fields produce no change in energy and it is therefore a reasonable assumption that they cannot cause an electron to bridge the energy gap between two bands.

When an external constant field $\mathbf{F}$ is applied the usual perturbation theory gives (Jones and Zener 1934 a)

$$\frac{h}{2\pi}\frac{\partial}{\partial t}|g(\mathbf{k}, t)|^2 = e\mathbf{F}.\operatorname{grad}_{\mathbf{k}}|g(\mathbf{k}, t)|^2. \qquad \ldots\ldots(10)$$

From this it follows that the effect of the field is to produce a uniform translation in time, parallel to the field, of the distribution in k-space, the analogy with the classical result being obvious.

In the same way the perturbing effect of a magnetic field $\mathbf{H}$ is given by

$$\frac{\partial}{\partial t}|g(\mathbf{k}, t)|^2 = -\frac{2\pi e}{hc}(\mathbf{H} \times \mathbf{v}_{\mathbf{k}}).\nabla_{\mathbf{k}}|g(\mathbf{k}, t)|^2, \qquad \ldots\ldots(11)$$

when terms in H^2 are neglected, which is permissible provided that $\tau \ll 2mc/eH$ (Jones and Zener 1934 a). When this restriction is fulfilled we can use a statistical treatment to determine the distribution function for electrons in $\mathbf{k}$-space, expressing by means of a Boltzmann equation the condition that they are in a steady state.

Let us consider then a model in which the conducting electrons all lie in a single energy band. In the absence of external fields if $E(\mathbf{k})$ is the energy of an electron in the state $\mathbf{k}$, the distribution of electrons in $\mathbf{k}$-space, at temperature T, will be just the Fermi distribution, that is, the number of electrons with momentum in the volume element dk_1, dk_2, dk_3 at (k_1, k_2, k_3) will be $(1/4\pi^3)f_0(\mathbf{k})\, dk_1\, dk_2\, dk_3$ where* $f_0(\mathbf{k}) = \{\exp[\{E(\mathbf{k}) - \zeta\}/\boldsymbol{k}T] + 1\}^{-1}$ and ζ is the chemical potential of the electron gas.

* The factor $\boldsymbol{k}$ in the denominator of the exponential is the usual Boltzmann constant and must not be confused with the vector $\mathbf{k}$.

If now the metal is placed in external electric and magnetic fields these will cause changes in the distribution function $f(\mathbf{k},t)=|g(\mathbf{k},t)|^2$, in the way expressed by equations (10) and (11), giving the total rate of change of $f(\mathbf{k},t)$ due to fields as

$$\left(\frac{\partial f}{\partial t}\right)_{\text{fields}} = +\frac{2\pi e}{h}(\mathbf{F}.\nabla_{\mathbf{k}})f(\mathbf{k},t) - \frac{2\pi e}{hc}(\mathbf{H}\times\mathbf{v}_{\mathbf{k}}).\nabla_{\mathbf{k}}f(\mathbf{k},t) \quad \ldots\ldots(12)$$

when temperature gradients and edge effects are neglected.

In order that the system shall be in an overall steady state $f(\mathbf{k},t)$ must in fact be independent of t, and the damping effect of the thermal vibrations and any impurities in the lattice must be just sufficient to balance the effects of the fields. The simplest assumption which we can make is that there exists a relaxation time τ which is independent of direction in $\mathbf{k}$-space (but may depend on $|\mathbf{k}|$), giving the rate of change of $f(\mathbf{k},t)$ as a result of collisions to be

$$\partial f/\partial t = -(f-f_0)/\tau. \quad \ldots\ldots(13)$$

More generally τ may be regarded as a function of $\mathbf{k}$. If in fact τ exists and *is* isotropic in k-space, and the lattice binding is also isotropic then we get no magneto-resistance effect, or rather, in the case of transverse magnetic fields a magneto-resistance effect (arising from thermal dispersion) which is far too small, and in the case of longitudinal fields no magneto-resistance effect whatsoever. Wilson (1936, p. 208) has shown that in the case of a metal in which the binding forces are isotropic and the scattering is supposed to be isotropic, an isotropic relaxation time exists only in the two extreme cases (*a*) of very high temperatures $T\gg\Theta$, where Θ is the Debye temperature and (*b*) at very low temperatures when the thermal vibrations are negligible as regards scattering power compared with that of the distortions produced by impurities in the lattice, and a relaxation time for the impurity scattering again exists.

For the general case we write then (cf. eqn. (13))

$$(\partial f/\partial t)_{\text{coll}} = \Phi f, \quad \ldots\ldots(14)$$

where Φ is a linear operator. Now we have

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = \iiint [W(\mathbf{k}',\mathbf{k})f(\mathbf{k}')\{1-f(\mathbf{k})\} - W(\mathbf{k},\mathbf{k}')f(\mathbf{k})\{1-f(\mathbf{k}')\}]\, dk_1'\, dk_2'\, dk_3' \quad \ldots\ldots(15)$$

where $W(\mathbf{k},\mathbf{k}')$ is the probability per unit time that an electron in a state k will make a transition to the state k'. Assuming that $E_{\mathbf{k}}$ is a function of k^2 only and that the thermal vibrations of the lattice are isotropic it is possible to reduce (15) to give (Bloch 1930)

$$\left(\frac{\partial f}{\partial t}\right)_{\text{coll}} = -\frac{(2m)^{3/2}}{4h^3\Lambda}\left(\frac{T}{\Theta}\right)^3 \int_{-\Theta/T}^{\Theta/T} [Ec(E) - c(E+\mathbf{k}Tz)\{E+\tfrac{1}{2}\mathbf{k}Tz - D(T/\Theta)^2z^2\}] \times \frac{e^{\eta+1}}{e^{\eta+z}+1}\frac{z^2\,dz}{|1-e^{-z}|} \quad \ldots\ldots(16)$$

when we assume the particular form of solution $f(\mathbf{k}) = f_0(\mathbf{k}) - k_1 c(E)\partial f_0/\partial E$. Λ and D are constants characteristic of the metal, Θ the Debye temperature and $\eta=(E-\zeta)/\mathbf{k}T$. The effect of impurities may be obtained by adding a term involving a relaxation time (Wilson 1937, eqn. (18)).

If we now write $f=f_0+f_1$ and assume that Ohm's law is universally true (that is that terms in F^2 can be neglected), the Boltzmann equation is

$$-\frac{2\pi e}{h}(\mathbf{F}.\nabla_{\mathbf{k}})f_0(\mathbf{k})+\frac{2\pi e}{hc}(\mathbf{H}\times\mathbf{v}_{\mathbf{k}}).\nabla_{\mathbf{k}}f_1(\mathbf{k})=\Phi f_1(\mathbf{k}). \quad \ldots\ldots(17)$$

It will be noted that in obtaining the expression (16) for $(\partial f/\partial t)_{\text{coll}}$ and in general treatments of Φ it is assumed that this collision process is independent of the applied fields. This will be true when the applied fields are small, but should they be large enough to make a considerable alteration in the electron states between collisions it is doubtful whether the different effects can be treated as independent. Bardeen and Shockley (1950) have shown that in the case of an applied electric field only it is valid to treat the thermal scattering as independent of the field if this is less than about 6×10^5 volts/cm which is generally the case.

§ 13. SOLUTION OF THE BOLTZMANN EQUATION AND DETERMINATION OF RESISTANCE

In general it is not possible to obtain an exact solution of the Boltzmann equation and we are led to consider such limiting cases as high and low magnetic fields or to assume the existence of an isotropic relaxation time. We must remember however that in the case of very high fields the physical validity of the Boltzmann equation is to be questioned and we must not then expect to get results which are necessarily quantitatively correct.

Let us consider the simplest model possible, namely that of a free electron gas in which $E=(h^2/8\pi^2\lambda)(k_1^2+k_2^2+k_3^2)$, where λ is the apparent mass of the electrons, and a relaxation time depending on $|\mathbf{k}|$ only exists; that is, it is a completely isotropic model. We find in this case that, when the electrons are assumed to be completely degenerate, i.e. $f_0(E)$ is treated as a step function which has the value 1 for $E<\zeta$ and 0 for $E>\zeta$, there is in fact no magneto-resistance effect (Sommerfeld and Frank 1931). Let us suppose that the current is in the x direction and the magnetic field perpendicular, in the z direction. The Boltzmann equation becomes

$$\frac{f_1}{\tau}+\frac{2\pi eH}{hc}\left(v_x\frac{\partial f_1}{\partial k_2}-v_y\frac{\partial f_1}{\partial k_1}\right)=\frac{eh}{2\pi\lambda}\frac{\partial f_0}{\partial E}(F_xk_1+F_yk_2) \quad \ldots\ldots(18)$$

where F_y is the Hall field required to prevent the flow of current in the y direction when the applied electric field is in the x direction. Since we are dealing with free electrons we have $\mathbf{v}_{\mathbf{k}}=(h/2\pi\lambda)\mathbf{k}$. Using this and multiplying through (18) by v_x and v_y in turn and integrating over k-space we get

$$\frac{1}{\tau}J_x+\frac{e}{c}\frac{H}{\lambda}J_y=-\frac{4\pi}{3}e^2v^2F_x \quad \ldots\ldots(19a)$$

$$\frac{1}{\tau}J_y-\frac{e}{c}\frac{H}{\lambda}J_x=-\frac{4\pi}{3}e^2v^2F_y \quad \ldots\ldots(19b)$$

where v is the magnitude of the velocity on the Fermi surface. The restriction that $J_y=0$ gives

$$F_y=-eH\tau F_x/c\lambda, \quad \ldots\ldots(20)$$

i.e. the Hall field, cf. eqn. (3), and then (19b) shows that J_x is independent of H, that is there is a complete absence of magneto-resistance. Sommerfeld and Frank were only able to obtain a non-zero effect by assuming a spread of velocity in the conducting electrons which they assumed arose from thermal energy, but in fact the effect they predict is far smaller than that obtained experimentally.

For the case of a magnetic field parallel to the applied electric field there is again no magneto-resistance effect in this model even when we assume non-degeneracy of the electron states. According to experiment the longitudinal and transverse effects should be of the same order of magnitude (although the longitudinal is generally smaller) and this again points to the fact that our model is inadequate.

It appears then that magneto-resistance as observed in metals is an outcome of either anisotropy of the binding forces of the lattice or anisotropy of the scattering effect of the thermal vibrations or of both of these simultaneously.* Anisotropy of the binding forces means that the surfaces of constant energy in k-space are no longer spheres, as in the case of a free electron gas, but take a quasi-ellipsoidal form. In such a circumstance the application of a magnetic field is bound to have some effect, for it may be seen from (11) that a magnetic field causes a rotation of states along surfaces of constant energy in $\mathbf{k}$-space. When these surfaces are spheres, no change is produced, but the moment they become non-spherical their orientation in space will be changed by the field. Anisotropy of the scattering of the lattice means that the collision operator Φ is a function of direction in $\mathbf{k}$-space: what this amounts to is that the lattice vibrations are not isotropic and consequently have differential scattering effect with direction.

By consideration of the solution of the Schrödinger equation for a metallic electron in the absence of fields it appears that the surfaces of equal energy are roughly spherical so long as we do not approach too closely the boundary of a Brillouin zone. Now let us for definiteness consider the alkali metals: in their case within each zone, allowing for spin, there are sufficient energy states to accommodate two free electrons per atom, while for these metals it seems most reasonable to assume that there is only approximately one free electron per atom. Consequently the first Brillouin zone (which is a dodecahedron) is only half filled, and a sphere containing only half the total number of k-states in the zone (and thus having half the volume of the zone) does not approach the zone boundary closely at any place. Thus for these metals it seems reasonable to assume isotropic binding. The case of lithium has in fact been worked out in some detail and supports this assumption (Jones and Zener 1934 b). On the other hand, the *elastic* properties of the alkali metals are not isotropic (Fuchs 1936 a, b) and this suggests that the thermal vibrations themselves are not isotropic and consequently the collision operator will depend directionally on $\mathbf{k}$. The magneto-resistance effect in these metals must thus be attributed to anisotropy of scattering.

Blochinzev and Nordheim (1933) have considered a model exhibiting anisotropy of binding forces, by assuming the Fermi surface to be composed of a number of parts each of which is a quadratic function of k_1, k_2, k_3, but the scattering is assumed to be isotropic, in fact $(\partial f/\partial t)_{\text{coll}}$ is assumed to be expressible in terms of an isotropic relaxation time τ. In order to compare the values of magneto-resistance which they obtain with experiment they have to assume rather large values for the curvature of the Fermi surface and this would indicate the existence of binding forces several hundred times as great as other evidence would lead one to expect. This then indicates that a considerable part of the magneto-resistance effect arises from anisotropy of τ: that is, the anisotropy of scattering.

* We believe that in fact it would not be possible for entire isotropy of one to exist with anisotropy of the other.

In the two extreme cases of small and high magnetic fields the Boltzmann equation may be solved to any degree of approximation required in ascending powers of H and $1/H$ respectively; although of course for large H we must bear in mind the physical limitations already discussed.

Consider then the case of a *small* magnetic field H parallel to the z axis, and assume the existence of a relaxation time $\tau(\mathbf{k})$ depending on $\mathbf{k}$; then the solution of the appropriate Boltzmann equation, which may be written

$$\frac{4\pi^2 eH}{h^2 c}\Omega c_1(\mathbf{k}) - \frac{1}{\tau(\mathbf{k})}c_1(\mathbf{k}) = \frac{2\pi e}{h}(\mathbf{F}.\operatorname{grad}_{\mathbf{k}} E),$$

with $$f_1(\mathbf{k}) = -\frac{\partial f_0}{\partial E}c_1(\mathbf{k}) \quad \text{and} \quad \Omega \equiv \frac{\partial E}{\partial k_2}\frac{\partial}{\partial k_1} - \frac{\partial E}{\partial k_1}\frac{\partial}{\partial k_2},$$

is (Jones and Zener 1934 b)

$$c_1(\mathbf{k}) = -\frac{2\pi e}{h}\left[\tau(\mathbf{F}.\operatorname{grad}_{\mathbf{k}} E) + \frac{4\pi^2 eH}{h^2 c}\tau\Omega(\tau\mathbf{F}.\operatorname{grad}_{\mathbf{k}} E) + \left(\frac{4\pi^2 eH}{h^2 c}\right)^2 \tau\Omega\{\tau\Omega(\tau\mathbf{F}.\operatorname{grad}_{\mathbf{k}} E)\}\right].$$

Taking now $$J = +\frac{2\pi e}{h}\int\int\int \operatorname{grad}_{\mathbf{k}} E \frac{\partial f_0}{\partial E}c_1(\mathbf{k})\, dk_1\, dk_2\, dk_3$$

we get immediately expressions for the current. It is readily shown quite generally from the resulting expressions that the magneto-resistance is non-negative in the case of both longitudinal and transverse magnetic fields. As we should expect from previous work the expressions give zero magneto-resistance when the energy surfaces are spherical, and τ isotropic, unless we take into account a departure of the electrons from complete degeneracy. Jones and Zener have obtained numerical results in the case of lithium making use of the fact that the energy surfaces are *almost* spherical. They obtain results of the correct order of magnitude but smaller than those obtained by experiment. This is to be expected for it is believed that in lithium the scattering also is not isotropic and consequently this will also make some contribution to the magneto-resistance.

In the case of *very high* magnetic fields the disturbing effect due to the field will be large compared with that due to the lattice vibrations. In the first approximation it would appear reasonable then to neglect the term $(\partial f/\partial t)_{\text{coll}}$ in the Boltzmann equation. This is equivalent to saying that approximately the Hall field is able to maintain a statistically steady state of the electrons, the collisions with lattice vibrations producing only a small second-order effect. By this means it is possible to find approximate solutions of the Boltzmann equation for the whole temperature range when the applied magnetic field is large. In the case of a pure isotropic metal and using the integral expression (16) for $(\partial f/\partial t)_{\text{coll}}$, Kohler (1940) has shown that in the presence of a large transverse magnetic field the usual interpolation formula for electrical conductivity (see, for example, Wilson 1936, p. 214) is

$$\sigma = K\left(\frac{\Theta}{T}\right)^5 \frac{1}{I_5},$$

where K is a constant depending on the metal, Θ is the Debye temperature and

$$I_5 = \int_0^\infty \frac{z^5\, dz}{(e^z-1)(1-e^{-z})},$$

is in fact exactly true when terms in $1/H^2$ are neglected. Thus the interpolation formula gives in fact the saturation value (*for the particular model employed*) attained

by resistance in the limiting case $H \to \infty$. Now at very high temperatures, since the collision integral (16) can then be expressed in terms of an isotropic relaxation time, and we have already shown that in such a case magneto-resistance is zero, it is hardly surprising that the two results coincide. Similarly with the model which we are using the magneto-resistance should also vanish at very low temperatures, since the approximation method adopted at very low temperatures for solving the Boltzmann equation and which leads to the formula for σ above, essentially (as far as the approximation goes) expresses $(\partial f/\partial t)_{\text{coll}}$ in terms of an isotropic relaxation time. In the *intermediate* temperature regions since magneto-resistance is essentially non-negative, we see that the usual interpolation formula as above gives an expression for resistance which is greater than the actual resistance (cf. recent papers by Kohler 1948, 1949a, Rhodes 1950, Sondheimer 1950 b).

If further we allow, in this same model, a term in the Boltzmann equation for impurity scattering, expressed in terms of a relaxation time, then it is found that Matthiessen's rule becomes strictly true. An examination of the analysis shows that this is hardly accidental, since the effect of high magnetic fields, when we solve the Boltzmann equation only to the power of $1/H^2$, is to *impose* an isotropic relaxation time on to the system and it is in fact the case that whenever a relaxation time exists, Matthiessen's rule holds. If we go to a higher approximation an isotropic relaxation time no longer exists and consequently we shall find deviations from Matthiessen's rule.

This approximation method of solving the Boltzmann equation is legitimate only so long as $\tau \gg mc/eH$, τ being of the order of magnitude of the relaxation time. But this inequality is just the direct negation of the condition that the Boltzmann equation shall be valid and consequently we should not expect the results to agree quantitatively with experiment.

The same method has been used by Justi and Kohler (1939) to discuss the effect of high magnetic fields on an alkali-type metal, that is a monovalent metal with cubic lattice structure, in which binding is supposed to be isotropic (i.e. $E \propto |\mathbf{k}|^2$), but scattering is assumed to be *anisotropic*. They assumed that $(\partial f/\partial t)_{\text{coll}}$ can be represented by a generalized relaxation time $\tau(\mathbf{k})$, which of course must have the same symmetry as the lattice. They are then able to show that this model will have a non-zero magneto-resistance which attains a saturation value for very high fields. (We may compare equation (6) of § 3.)

It is in fact the case that experiment indicates two contrasting modes of behaviour of a metal when high magnetic fields are applied. In some, on the one hand, the magneto-resistance tends towards a saturation value as the field becomes very large (see, for example, Justi and Scheffers 1938); in others (cf. Justi and Scheffers 1936, 1937), it appears ultimately to increase about proportionally to H^2 (cf. also fig. 4). The odd-valent (in particular, monovalent) metals appear to follow the former characteristic, whereas the second is shown by even-valent (in particular, divalent) metals. Now monovalent metals have either the face-centred or body-centred cubic lattice structure, and one free conducting electron per atom. In such a case, assuming degeneracy, the Fermi surface lies well within the first Brillouin zone and consequently we need only consider the electrons as lying in one zone. On the other hand, in a divalent metal the electrons penetrate into a second zone. In such a case we should have contributions to the current from the electrons in the upper zone and the holes in the lower zone.

With such a model it is possible to get an increase of magneto-resistance proportional to H^2 in large fields provided that the number of holes in one band is equal to the number of electrons in the other. The two cases have been considered in detail by Kohler (1949a, b). Let us consider first the case of a monovalent metal having the properties described, taking the coordinate axes O_x, O_y, O_z parallel to the principal crystal axes, and the magnetic field H parallel to Oz. The Boltzmann equation (17) can be written

$$-e(\mathbf{F}.\mathbf{v})\frac{\partial f_0}{\partial E}+\frac{2\pi eH}{hc}\left(v_x\frac{\partial f_1}{\partial k_2}-v_y\frac{\partial f_1}{\partial k_1}\right)=\Phi f_1(\mathbf{k}).$$

When we neglect the collision term on the right-hand side, the resulting equation has the solution, to this approximation,

$$f_1'=\frac{hc}{2\pi}\frac{1}{H}[F_xk_2-F_yk_1],$$

and we see that in this approximation F_z must be zero, that is the disturbance produced by F_z can be at most as great as that produced by the collisions. In this approximation the current has components

$$J_x'=-\frac{e}{c}\frac{1}{H}\iiint(F_xk_2-F_yk_1)\frac{\partial E}{\partial k_1}\frac{\partial f_0}{\partial E}d\tau_{\mathbf{k}}=+\frac{e}{c}\frac{1}{H}t_{11}F_y$$

$$J_y'=-\frac{e}{c}\frac{1}{H}\iiint(F_xk_2-F_yk_1)\frac{\partial E}{\partial k_2}\frac{\partial f_0}{\partial E}d\tau_{\mathbf{k}}=-\frac{e}{c}\frac{1}{H}t_{11}F_x$$

$$J_z'=0$$

where t_{pq} is a tensor, equal to $\iiint k_p\frac{\partial E}{\partial k_q}d\tau_{\mathbf{k}}$, which since it has the symmetry of the lattice must satisfy $t_{pq}=0$ $p\neq q$, $t_{11}=t_{22}=t_{33}$.

To determine the second approximation to the distribution function, we write $f_1=f_1'+f_1''+\ldots.$, and f_1'' satisfies

$$-eF_xv_x\frac{\partial f_0}{\partial E}+\frac{2\pi e}{hc}H\left(v_x\frac{\partial f_1''}{\partial k_2}-v_y\frac{\partial f_1''}{\partial k_1}\right)=\Phi f_1'(\mathbf{k}) \qquad \ldots\ldots(21)$$

where F_x must be $O(1/H)$ and f_1'' is $O(1/H^2)$. Multiplying through (21) by k_y and integrating over the first Brillouin zone gives, after some reduction, when we write $J_x=J_x'+J_x''+\ldots.$

$$J_x''=\left(\frac{hc}{H}\right)^2\Phi_{22}F_x-\frac{e}{H}\iint(\mathbf{v}\times\mathbf{H})_\nu.f_2k_y\,d\sigma$$

where $\phi_{pq}=\iint k_p\Phi k_q\,d\tau_{\mathbf{k}}$ and is such that $\phi_{pp}>0$ (Kohler 1942) and for a crystal with cubic symmetry $\phi_{11}=\phi_{22}=\phi_{33}$. The surface integral is taken over the boundary of the Brillouin zone, the subscript ν of the vector product denoting the component of a vector along the outward normal to the surface.

The expression for J_y'' is similar. We then have for the conductivity

$$\sigma=\frac{J_x^2+J_y^2+J_z^2}{F_xJ_x+F_yJ_y+F_zJ_z}$$

$$=\frac{e^2}{h^2}\frac{(F_x^2+F_y^2)t_{pp}^2}{\phi_{pp}(F_x^2+F_y^2)-eH/(ch)^2\iint(\mathbf{v}\times\mathbf{H})f_2(F_xk_y-F_yk_x)\,d\sigma}, \qquad \ldots\ldots(22)$$

neglecting terms in $1/H^2$ in the numerator and the denominator (after cancelling by a factor $1/H^2$). Since f_2 is $O(1/H^2)$, the conductivity shows saturation,

provided of course that F_x and F_y are not both zero. Now in the case we are considering, the Fermi surface is assumed to be well within the first Brillouin zone and consequently the surface integral appearing in σ will vanish since the factor f_3 contains $\partial f_0/dE$, so that in this case $\sigma=(e^2/h^2)t_{pp}/\phi_{pp}$ and is isotropic. It is in fact true that most metals show some anisotropy in magneto-resistance (this is true even in sodium, but much more strongly in other metals, for example, gold). According to this theory such anisotropy is to be explained by the non-vanishing of the surface integral in the equation for J_x'', which is equivalent to saying that the Fermi surface approaches the boundary of the zone within a distance of the order of magnitude of $\boldsymbol{k}T$.

The method breaks down when we have a longitudinal magnetic field. In this case by assuming a solution of the form $f_1=(\partial f_0/\partial E)g(E)k_3$ where

$$g(E)=\Sigma_{m=0}^{\infty}A_m(E-\zeta)^m,$$

Kohler finds as a first approximation that the resistance is $(h^2/e^2)\phi_{pp}/t_{pp}$, i.e. the same as in the transverse case when this is isotropic. Higher approximations give a smaller value for the longitudinal resistance. Thus we have the result that if the Fermi surface does not approach the boundary of the first Brillouin zone the longitudinal magneto-resistance is less than or equal to the transverse magneto-resistance.

In the case of a divalent metal, let us consider a magnetic field in the z direction and the *resultant* electric field in the x direction. Then in the first approximation, for a metal with cubic lattice structure, we have as above: $J_x'=J_z'=0$, $J_y'=(e/cH)t_{pp}F_x$.

The conductivity is now given by

$$\sigma_H=\frac{(J_x''+J_x'''+\ldots.)^2+(J_y'+J_y''+\ldots.)^2+(J_z'+\ldots.)^2}{F_x(J_x''+J_x'''\ldots.)}.$$

Unless J_y' is zero the numerator and denominator are both of order $1/H^2$ and there will be saturation. Only when J_y' vanishes can the conductivity decrease as $1/H^2$, and this requires that t_{pp} shall be zero. When all the electrons are contained within the first zone, the volume integral which gives t_{pp} has to be taken over this zone only, and Kohler shows, by using Gauss' theorem, that it is just equal to the number of 'holes' in the zone. This is zero in the case of a monovalent metal. In the case of a divalent metal however the electrons penetrate into the second Brillouin zone and the integration for t_{pp} is this case has to extend over the first two zones. Again the contribution from the first zone is equal to the number of holes in the zone, while the contribution from the second is equal to minus the number of electrons in the zone. Thus when the number of holes in the first zone is equal to the number of electrons in the second, $t_{pp}=0$, and in this case we shall have an H^2 law for resistance in high magnetic fields. By a similar method it is possible to show that the same law for increase of resistance holds good for crystals with more complicated crystal structure than cubic.

The same result was obtained by Blochinzev and Nordheim (1933) in discussing a divalent metal as well as by Sondheimer and Wilson, who also worked with a two-band model. Sondheimer and Wilson assume energy surfaces with spherical symmetry and such that there is an overlap of energy between the two bands. They neglect any transitions between the two bands and thus assume that each contributes separately to the current. In the two limiting cases of high temperatures,

when an isotropic relaxation time exists, and of low temperatures where an approximation method can be adopted, the Boltzmann equation can be solved in the usual way. By expressing the relative change of resistance at these extremes of temperature in terms of the resistance ρ_0 in the *absence* of magnetic field, it is suggested that if we then interpret ρ_0 to be the usual interpolation expression, including a term for impurity scattering (Wilson 1937),

$$\rho_0 = \frac{3\pi^2\hbar^4\beta}{(2m)^{3/2}e^2\zeta} + \frac{3\pi^2\hbar^2 D}{2e^2\zeta^2\Lambda}\left(\frac{T}{\Theta}\right)^5 I_5,$$

where D and Λ are specific constants of the metal, the resulting expressions for magneto-resistance will be valid over the whole temperature range. Only in the special case when the number of electrons in one band n_s is equal to the number of holes in the other n_d is there an increase in resistance proportional to H^2 for large fields, and in fact this law holds for all field strengths. Otherwise there is the usual increase proportional to H^2 for small fields, and ultimately saturation for very large fields with an increase slower than H^2 for intermediate field strengths. As regards the temperature variation, with a *constant applied magnetic field* when $n_s = n_d$, or the difference between n_s and n_d is small, for small values of H ($< nec\rho_r$, ρ_r = residual resistance, n = number of free electrons) the resistance increases steadily with T, while for $H > nec\rho_r$ the resistance first *decreases* with increasing T to a minimum and then increases. When the difference between n_s and n_d is sufficiently large the resistance increases steadily with T for all H.

§ 14. SIZE EFFECT

In a discussion of the effect of the size of a specimen on magneto-resistance the Boltzmann equation has to be modified to allow for the fact that the electron distribution depends on *position* coordinates as well as $\mathbf{k}$ coordinates. Let us suppose that the specimen is in the shape of a plate of uniform (small) thickness, its dimensions in the other two perpendicular directions being so large that they can be considered as infinite. Then if the z axis is taken perpendicular to the plane faces of the plate the distribution function will depend on $\mathbf{k}$ and z. From the consideration that as many electrons must enter a strip of thickness δz parallel to the plane faces it will be seen that the term $v_z(\partial f/\partial z)$ must be added to the expression (12) for $(\partial f/\partial t)_{\text{fields}}$.

Assuming the specimen otherwise isotropic, i.e. as regards thermal and impurity scattering and binding forces, various cases corresponding to differing directions of the magnetic field have been discussed (Sondheimer 1950 a, Chambers 1950, MacDonald and Sarginson 1950; cf. also Dingle (1950)). They all show an early fall in resistance when a magnetic field is applied and thus support the experimental results of MacDonald (MacDonald 1949, MacDonald and Sarginson 1950). The case in which both the magnetic field and the applied electric field are in the plane of the plate has some interest apart from the discussion of the anomalous behaviour of magneto-resistance, for in this case a charge density is set up at the edges of the plate and consequently the Hall field is *not* constant across the metal. The Hall field is in fact given by what is, in general, a rather cumbersome integral equation. In the case of a small applied magnetic field an approximation to this equation can be used and the solution indicates a Hall field which rises rather rapidly at the edge of the plate, to something of the order of magnitude of its bulk value.

REFERENCES

BARDEEN, J., and SHOCKLEY, W., 1950, *Phys. Rev.*, **80**, 69.

BETHE, H., 1931, *Nature, Lond.*, **127**, 336.

BLACKMAN, M., 1951, *Proc. Phys. Soc.* A, **64**, 681.

BLOCH, F., 1929, *Z. Phys.*, **52**, 555; 1930, *Ibid.*, **59**, 208.

BLOCHINZEV, D., and NORDHEIM, L., 1933, *Z. Phys.*, **84**, 168.

BLOM, J. W., 1950, *Physica*, **16**, 144, 152, 171.

BUSCH, G., 1951, *Semi-conducting Materials*, ed. H. K. Henisch (London : Butterworths Scientific Publications).

BUSCH, G., WIELAND, J., and ZOLLER, H., 1951, *Helv. phys. Acta*, **24**, 49.

CHAMBERS, R. G., 1950, *Proc. Roy. Soc.* A, **202**, 378.

CONWELL, E., and WEISSKOPF, V. F., 1950, *Phys. Rev.*, **77**, 388.

CRAWFORD, J. H., and LARK-HOROVITZ, K., 1950, *Phys. Rev.*, **79**, 889.

CORNISH, F., and MACDONALD, D. K. C., 1951, *Phil. Mag.*, **42**, 1406.

DINGLE, R. B., 1950, *Proc. Roy. Soc.* A, **201**, 545.

ESHELBY, J. D., FRANK, F. C., and NABARRO, F. R. N., 1951, *Phil. Mag.*, **42**, 351.

FRANK, F. C., 1951, *Phil. Mag.*, **42**, 809.

FUCHS, K., 1936 a, *Proc. Roy. Soc.* A, **153**, 622; 1936 b, *Ibid.*, **157**, 444; 1938, *Proc. Camb. Phil. Soc.*, **34**, 100.

GERRITSEN, A. N., and LINDE, J. O., 1951, *Physica*, **17**, 573, 584.

HALL, E. H., 1879, *Amer. J. Math.* (or see *Phil. Mag.*, 1880, **9**, 225); 1880, *Phil. Mag.*, **10**, 301; 1881, *Ibid.*, **12**, 157.

JOHNSON, V. A., and LARK-HOROVITZ, K., 1950, *Phys. Rev.*, **79**, 176, 409.

JONES, H., 1951, *Phys. Rev.*, **81**, 149.

JONES, H., and ZENER, C., 1934 a, *Proc. Roy. Soc.* A, **144**, 101; 1934 b, *Ibid.*, **145**, 268.

JUSTI, E., 1940, *Z. tech. Phys.*, **21**, 315; 1948, *Leitfähigkeit und Leitungsmechanismus fester Stoffe* (Göttingen : Vandenhoeck and Ruprecht).

JUSTI, E., and KOHLER, M., 1939, *Ann. Phys., Lpz.*, **36**, 349.

JUSTI, E., and SCHEFFERS, H., 1936, *Phys. Z.*, **37**, 700; 1937, *Ibid.*, **38**, 891; 1938, *Ibid.*, **39**, 105.

KAPITZA, P., 1927, *Proc. Roy. Soc.* A, **115**, 658; 1928, *Ibid.*, **119**, 358; 1929, *Ibid.*, **123**, 292.

KELVIN, LORD, 1856, *Mathematics and Physics Papers*, **2**, 307.

KLAHR, C. N., 1951, *Phys. Rev.*, **82**, 109.

KOEHLER, J. S., 1942, *Amer. J. Phys.*, **10**, 275; 1949, *Phys. Rev.*, **75**, 106.

KOHLER, M., 1938, *Phys. Z.*, **39**, 9; 1940, *Ann. Phys., Lpz.*, **38**, 283; 1942, *Ibid.*, **42**, 142; 1948, *Z. Phys.*, **124**, 772; 1949 a, *Ibid*, **125**, 679; 1949 b, *Ann. Phys., Lpz.*, **5**, 89; 1949 c, *Ibid.*, **5**, 99.

KOMAR, A., 1941, *J. Phys., U.S.S.R.*, **4**, 547.

KOMAR, A., and SIDOROV, S., 1941, *J. Phys., U.S.S.R.*, **4**, 552.

KUNDT, A., 1893, *Wied. Annal.*, **49**, 257.

MACDONALD, D. K. C., 1949, *Nature, Lond.*, **163**, 637; 1950, *Proc. Phys. Soc.* A, **63**, 290; 1951, *Phil. Mag.*, **42**, 756.

MACDONALD, D. K. C., and MENDELSSOHN, K., 1948, *Nature, Lond.*, **161**, 972; 1950, *Proc. Roy. Soc.* A, **202**, 103.

MACDONALD, D. K. C., and SARGINSON, K., 1950, *Proc. Roy. Soc.* A, **203**, 223.

MACKENZIE, J. K., and SONDHEIMER, E. H., 1950, *Phys. Rev.*, **77**, 264.

MENDOZA, E., and THOMAS, J. G., 1951, *Phil. Mag.*, **42**, 291.

MEISSNER, W., and SCHEFFERS, H., 1929, *Phys. Z.*, **30**, 827.

MILNER, C. J., 1937, *Proc. Roy. Soc.* A, **160**, 207.

MOTT, N. F., 1935, *Proc. Phys. Soc.*, **47**, 571; 1936, *Proc. Roy. Soc.* A, **153**, 699.

MOTT, N. F., and JONES, H., 1936, *Theory of Properties of Metals and Alloys* (Oxford : University Press).

NERNST, W., and v. ETTINGHAUSEN, A., 1886, *Wied. Annal.*, **29**, 343.

PEARSON, G. L., HAYNES, J. R., and SHOCKLEY, W., 1950, *Phys. Rev.*, **78**, 295.

PEIERLS, R., 1931, *Ann. Phys., Lpz.*, **10**, 97; 1933, *Z. Phys.*, **80**, 763.

PUGH, E. M., ROSTOKER, N., and SCHINDLER, A., 1950, *Phys. Rev.*, **80**, 688.

RHODES, P., 1950, *Proc. Roy. Soc.* A, **202**, 466.

ROSTOKER, N., and PUGH, E. M., 1951, *Phys. Rev.*, **82**, 125.

SEITZ, F., 1940, *The Modern Theory of Solids* (New York and London : McGraw Hill).
SKINNER, H. W. B., 1940, *Phil. Trans. Roy. Soc.* A, **239,** 95.
SMIT, J., 1951, *Physica*, **17,** 612.
SOMMERFELD, A., and BETHE, H., 1933, *Handbuch der Physik*, **24,** (2), 363 ff., 562 ff. (Berlin : Springer).
SOMMERFELD, A., and FRANK, N. H., 1931, *Rev. Mod. Phys.*, **3,** 1.
SONDHEIMER, E. H., 1950 a, *Phys. Rev.*, **80,** 401; 1950 b, *Proc. Roy. Soc.* A, **203,** 75.
SONDHEIMER, E. H., and WILSON, A. H., 1947, *Proc. Roy. Soc.* A, **190,** 435.
THOMSON, J. J. 1901, *Proc. Camb. Phil. Soc.*, **11,** 120.
ŢIŢEICA, Ş., 1935, *Ann. Phys., Lpz.*, **22,** 129.
TOMLINSON, H., 1883, *Phil. Trans. Roy. Soc.* A, **174,** 1. (See also Righi, 1884, *J. Phys. Radium*, **3,** 355.)
WILSON, A. H., 1936, *The Theory of Metals* (Cambridge : University Press); 1937, *Proc. Camb. Phil. Soc.*, **33,** 371.
ZENER, C., 1934, *Proc. Roy. Soc.* A, **145,** 523.

Although not referred to in the text we should also like to note :

BRONSTEIN, M., 1932, *Phys. Z. Sowjet*, **2,** 28 (On the Theory of Electronic Semiconductors).
FRÖHLICH, H. H. 1936, *Elektronentheorie der Metalle*, 212, 224 (Berlin : Springer).
MEISSNER, W., 1935, *Handbuch d. Exp. Phys.* vol. 11/2, Elektronenleitung. Galvanomagnetische, thermoelektrische und verwandte Effekte. (Leipzig : Akademie Verlag).

TRAVELLING-WAVE TUBES

By R. KOMPFNER

CONTENTS

	PAGE
§ 1. Introduction	276
§ 2. Historical background	276
§ 3. Early investigations leading to the discovery of the travelling-wave interaction	279
§ 4. Description of mechanism of travelling-wave tube gain	286
§ 5. Interim review of the problem	287
§ 6. Developments at the Bell Telephone Laboratories	288
§ 7. Theory of the travelling-wave tube according to Pierce	290
7.1. Basic equations	290
7.2. Propagation neglecting space charge and circuit attenuation; synchronous velocities	291
7.3. Propagation taking space charge and circuit attenuation into account; non-synchronous velocities	293
7.4. Power output and efficiency	296
7.5. The helix	296
§ 8. Further developments of the helix-type travelling-wave tube	300
8.1. Localized attenuation	300
8.2. Optimum proportions of helix	300
8.3. Dispersive travelling-wave tubes	301
8.4. Low-noise travelling-wave tubes	302
8.5. Space-charge smoothing at microwave frequencies	303
8.6. Electron optics in the travelling-wave tube	304
8.7. The helix-type travelling-wave tube carried to its limits	306
§ 9. Structures other than helical which support slow waves	307
9.1. Fundamental mode of operation	307
9.2. Harmonic mode of operation	309
§ 10. Travelling-wave interaction in crossed electric and magnetic fields	312
§ 11. Electron plasma waves	314
§ 12. The electron-wave tube	316
12.1. Theory	316
12.2. Experimental	318
§ 13. The slipping-stream tube	319
§ 14. The velocity-jump tube	320
§ 15. Other forms of travelling-wave tubes	321
15.1. Transverse current travelling-wave tube	321
15.2. The 'Easitron'	323
§ 16. Applications of the travelling-wave tube	324
References	325

ABSTRACT. 'Travelling-wave tubes' is the name of a recently developed class of electron tubes, operating in the microwave range of frequencies, which amplify by virtue of continuous interaction between an electron stream and a slow-travelling electromagnetic wave. The wave may be propagated along a retarding structure such as a helix or a corrugated waveguide, or it may be a space-charge wave in a second electron stream. The history of the discovery and the development of the travelling-wave tube is traced and a simple theory of operation outlined. A more complete theory according to J. R. Pierce is sketched by which can be calculated the effects of attenuation, of non-synchronous velocities and of space-charge effects. Recent developments relevant to the helix-type travelling-wave tube are described, such as 'Brillouin'-focusing of electron streams in axial magnetic fields, or noise properties of electron streams.

Other forms of travelling-wave tubes described include Millman's space-harmonic type of tube operating as amplifier at a wavelength of 6 millimetres, the electron-wave tube, and the travelling-wave magnetron amplifier which combines the high efficiency of the magnetron with its amplifying function.

Wherever possible attempts have been made to give explanations of the modes of operation of the various devices in simple physical terms.

§1. INTRODUCTION

MAJOR advances in pure science are often brought about by minor advances in applied science or technology. There can be little doubt for instance that the invention of electron tubes such as the triode, the photocell and the cathode-ray tube has enabled physical knowledge to be enriched to an extent perhaps only to be compared with the introduction of the microscope to biological science.

The importance of electron tubes to modern pure science is beyond doubt; many recently flourishing branches of science, such as microwave spectroscopy or radioastronomy, owe their very existence to advances made in the techniques of electron tubes or, as commonly and loosely called, 'electronics'. There is another aspect of this matter: it appears that processes discovered and studied in electron tubes may give clues to the mechanism of processes going on in the sun, and elsewhere in the universe, and thus studies in laboratories may lead to a better understanding of phenomena such as cosmic rays or cosmic radio noise. These matters are, however, still very much subjects under discussion at the time of writing.

Finally, the subject of this report is interesting in itself.

In this report we are concerned with a branch of the relatively new branch of electronics called 'microwave electronics'. Development in this branch is still going on at a furious pace, with new possibilities opening up all the time. The first practical result has appeared in the field of communication engineering.* No direct applications to pure science have been reported so far, although there is little doubt that such will appear in due course. Efforts are being made in many laboratories to study the implications of the newly discovered techniques in relation to other branches of technology, to extend the techniques in all possible directions and to find new fields in which to apply them.

Under these circumstances the historical method of reporting is favoured, although it makes it difficult to put things in proper perspective and in order of importance. Thus explained, it is hoped that the place of the account of the author's own investigations, which forms the beginning of this report, will be understood.

§2. HISTORICAL BACKGROUND

Generation and amplification of electromagnetic waves is nowadays commonly achieved by means of electron tubes. The most important example of such a tube is the triode, in which we can picture the process of conversion of d.c. power into radio-frequency power as follows: the input r.f. power produces an oscillatory electric field in the space between the cathode and the grid. This field modulates the intensity of the electron stream emitted from the cathode, that is to say, it superimposes an r.f. component on the 'd.c.' current. The modulated electron stream then enters the grid–anode space. An oscillatory electric field there, of the same frequency as the input r.f. power and in the right phase, will abstract net energy from the electron stream, inasmuch as more electrons arrive during the retarding phase of the field than during the accelerating phase. Part of the energy abstracted from the electron stream is available as r.f. output power and can be made to approach the available d.c. power, i.e. electron current times

* The B.B.C. television microwave link between Manchester and Scotland uses travelling-wave tubes.

cathode–anode potential, in magnitude. In principle, the triode can be a very efficient converter of d.c. into r.f. energy.

The process outlined above fails when the time taken by an electron to cross the cathode–grid interaction space becomes an appreciable fraction of a cycle. This is so not only because the electron density modulation is reduced, but also because now work has to be done by the r.f. field in the process. Similar considerations apply to the grid–anode interaction space, although matters are more favourable as the electrons there move considerably faster and transit times are shorter.

An important step in overcoming the transit-time limitation was taken when the so-called velocity-modulation principle was adopted (Heil and Heil 1935, Varian and Varian 1939, Hahn and Metcalf 1939). The simplest embodiment of this principle is the so-called klystron*.

In this electron tube the electron stream is first accelerated to a relatively high potential before it enters the input interaction space, usually called the 'buncher'. There an r.f. velocity modulation is impressed upon the beam, that is to say, electrons are accelerated during one part of the cycle and retarded during the other. The beam then passes into the so-called drift space, which is free from r.f. fields. Here electrons which entered later but which had been accelerated catch up on electrons which entered earlier but which had been retarded by the input r.f. field. (It is interesting to note that, again, this process absorbs no energy from the r.f. field provided electrons traverse the interaction space in a negligible fraction of a cycle.) At a suitable point along the drift space, where the bunching of the electrons is most favourable, i.e. where the r.f. component of current is a maximum, a second interaction space is provided, usually called the 'catcher'. When 'bunches' of current traverse the catcher field they induce fields there in such a phase that more electrons are slowed down than speeded up and r.f. energy is abstracted from the beam; part of this energy becomes available as r.f. output power.

Thus velocity modulation makes it possible to produce an intensity modulated electron stream by means of an interaction space which is traversed by relatively fast electrons. This is one of the main reasons why, at high frequencies, the klystron is superior to the triode. In the triode the electron stream enters the input interaction space with practically zero velocity, and therefore transit time is a far more important factor there at frequencies above, say, 1 000 Mc/s.

The same difficulties which set a limit to an efficient performance of the triode reappear when the klystron is used at still higher frequencies, say above 3 000 Mc/s. As the time taken by electrons to cross the input or output interaction spaces becomes an appreciable fraction of a cycle, the magnitude of the interaction decreases, while energy is now taken from the r.f. fields in the process of velocity modulation.

These limitations disappear altogether once use is made of the principle of *interaction between electron streams and travelling electromagnetic fields.*

The speed u_0 of an electron which has been accelerated to a potential V_0 is in the non-relativistic region given by $u_0 = (eV_0/m)^{1/2}$, where e and m are electron charge and mass respectively. In centimetres per second and volts we have $u_0 \simeq 6 \times 10^7 V_0^{1/2}$. Thus a potential of about 2 500 volts corresponds to an electron

* From the Greek 'κλυζειν', meaning the breaking of waves on the beach.

velocity of 3×10^9 cm sec^{-1}, which is about one-tenth of c, the velocity of propagation of electromagnetic waves in free space.

To produce travelling fields, therefore, which interact with electron streams accelerated by voltages conveniently obtained and handled in practice, that is, between a few hundred and a few thousand volts, means have to be found to reduce the effective velocity of electromagnetic waves drastically. There are two distinct methods which achieve this. The first, and simplest, although only employed comparatively recently, relies on direct slowing-down of a wave, such as occurs in a medium in which the dielectric constant or the permeability, or both, are larger than in free space. If it were possible to project an electron through such a medium with a velocity equal to that of the phase velocity of the wave, such an electron would remain in the same phase of the wave for a long time and thus experience a force substantially constant with time. The second method relies on the modification of a wave by a material periodic structure which produces a variation of the intensity of the wave periodic with distance. This periodic distance is not to be confused with the wavelength of the wave which is normally much larger. In a simple, hypothetical, case the intensity might be zero over a given distance, finite over a subsequent distance, zero again, finite again, and so on. Where it is finite, the intensity varies periodically in time with the frequency of the wave. Thus, if an electron travels from one location where the wave is finite to the next, in which it is finite again, in a time such that it finds the same phase of the field, the effect will be very much the same as if the electron had been in a field moving with it at the same speed, although the force on the electron now has a component periodic in time. If the wave travels substantially the distance of a whole wavelength while the electron moves from one interaction region to the next, the electron is said to interact with the first space harmonic (or Hartree harmonic) of the wave; if the wave moves two whole wavelengths we have the second space harmonic, and so on.

Historically, the second type of electromagnetic wave was the first travelling wave to be used for the purpose of acting on fast-moving charged particles (Wideröe 1928, Sloan and Lawrence 1931).

The first, and the most important, *electron tube*, based on the principle of interaction between electromagnetic fields and electrons travelling at about the same speed, which functioned in actual practice, is the multi-cavity magnetron generating microwave power at very high level and efficiency, invented and developed early during the 1939–1945 war (Boot and Randall 1946). The fact that the cavity magnetron operates on the travelling wave principle was not at first recognized. This, however, did not interfere with its success. We now realize that the multi-resonator cavity, i.e. the anode block, can be looked on as a wave-guiding structure capable of propagating a certain number of waves of different speeds, each wave furthermore consisting of an infinite number of space harmonics (Hartree harmonics). When a magnetron oscillates, the electron cloud emitted from the cathode rotates in the space between the cathode and anode with a velocity close to that of one of the possible waves and gives up energy to it. This process can be very efficient since electrons that travel in a phase of the wave in which they give up energy move progressively closer to the anode, without falling out of step with the wave, while those electrons that absorb energy from the wave are returned to the cathode at once and play no further part in the process.

In this report we shall not be concerned with the magnetron as oscillator, and for further information the reader is referred to the works of Fisk, Hagstrum and Hartman (1946) and Collins (1948). The magnetron principle as applied to amplification will be dealt with in the later sections of this report.

During the 1939–1945 war attempts were made in several places to use electron tubes as r.f. amplifiers to improve the sensitivity of radar receivers. These were successful at the lower frequencies in use earlier on, but failed at microwave frequencies. Work was done with triodes, tetrodes and klystrons (Foster 1946, Neher 1943, 1945, Kompfner 1947), and it appeared that it would be very difficult even to approach the noise figures of the crystal mixer–intermediate-frequency amplifier combination in common use. In every case theory showed that only low voltage tubes (100 volts or less) had any chance of success; this meant very close gap spacings if the transit times were to be kept low, and led to insuperable technical difficulties. Increasing the 'quality' of the circuits used with these tubes by making them more sharply resonant merely made them too selective, and reduced the frequency band of response below the minimum necessary for the reception of pulses as short as are used with radar.

At this stage it was natural that enquiries should be made to discover other means of obtaining useful interaction between electrons and r.f. fields. It must have occurred to many people that interaction between electrons and fields moving at the same speed as the electrons, i.e. travelling waves, offered better promise, and the writer was fortunate to have the opportunity and facilities to engage in some of the early researches.

It is possible now to see that a number of investigations, some earlier, some later, would eventually have led to the same results as were obtained by the writer. However, for the sake of clarity and continuity the first sections of this account of the early development of the travelling-wave tube will be based on published papers by the writer.

§3 EARLY INVESTIGATIONS LEADING TO THE DISCOVERY OF THE TRAVELLING-WAVE INTERACTION

If the physical extension of the stationary r.f. fields formed the basic difficulty with conventional electron tubes, it was argued that this could be overcome by making the field move with the electrons, and the extension of the field would turn from a liability into an asset. Extended fields might not be as intense as concentrated ones (where intensity is obtained by means of resonance), and to obtain comparable results interaction between the moving field and electron should be provided over long distances.

The first, and most important, step in this direction was to find a simple structure capable of slowing down a wave sufficiently and presenting the wave in a form well adapted for interaction with a long cylindrical electron beam. The first, and obvious, choice is a waveguide partly filled with a material of high dielectric constant and propagating a transverse magnetic wave. This was rejected because it was feared that such a material might become charged by electrons intercepted from the beam; moreover, a material with sufficiently high dielectric constant was not available at the time.

A helical transmission line was then considered. Such a transmission line consists of an inner conductor coiled in the form of a helix inside a solid cylindrical

outer conductor. It was expected that the increase in inductance of the inner conductor would produce some slowing down of a wave. Preliminary experiments showed that a slow wave could indeed be propagated along the helix with a phase velocity given over a wide range approximately by

$$v \sim \mathbf{c} \sin \psi, \qquad \ldots\ldots(3.1)$$

where ψ is the pitch angle of the helix. The attenuation was measured and found to be around a few decibels per metre. The impedance was estimated to be a few hundred ohms. Thus these experiments showed that a helical transmission line could indeed propagate a wave with a velocity corresponding to that of an electron beam accelerated to a potential around 1 000 volts, and crude calculations were made of the performances to be expected of tubes using helical transmission lines.

At first both transverse and axial fields seemed of interest. It was believed that deflection modulation of an electron beam resulting from interaction with transverse fields might give a better low-noise amplifier than axial interaction, i.e. velocity modulation. It was recognized that the wave would be strongest near the wire, decreasing with distance from the wire in a quasi-exponential manner. However, it was not known how quickly the field would fall off, and therefore 'weakening factors' (τ in the transverse, α in the axial case) were introduced. Assuming that a power P flowing in the transmission line gives rise to a peak voltage V on the helix, we have

$$P = V^2/2Z, \qquad \ldots\ldots(3.2)$$

where Z is the effective impedance of the line. Now we can define the effective transverse field between helix and outer conductor by

$$E_T = \tau V/D, \qquad \ldots\ldots(3.3)$$

where D is the distance between helix and outer conductor, and the effective axial field on the axis of the helix by

$$E_z = \alpha \partial V/\partial z = \alpha 2\pi V/\lambda, \qquad \ldots\ldots(3.4)$$

where λ is the wavelength along the helix.

It can then easily be shown that an electron travelling a distance z with a velocity u_0 in a transverse field E_T travelling with a phase velocity v will be displaced from its undisturbed position with a peak amplitude

$$s = \frac{\mathbf{e}E_T z^2}{2\mathbf{m}u_0^2}[(\sin\theta - \theta)^2 + (1 - \cos\theta)^2]^{1/2}\theta^{-2}, \qquad \ldots\ldots(3.5)$$

where θ is the phase shift in radians of an electron with respect to the wave, i.e.

$$\theta = \frac{2\pi z}{\lambda}\left(\frac{v}{u_0} - 1\right). \qquad \ldots\ldots(3.6)$$

When $u_0 = v$, $\theta \to 0$, (3.5) reduces to

$$s = \tau V z^2/4V_0 D, \qquad \ldots\ldots(3.5a)$$

which will be recognized as the displacement of electrons travelling through ordinary deflecting plates D distant and z long in the low-frequency case, i.e. when the transit time through the plates is a small fraction of a cycle. If ordinary deflecting plates of length comparable with, or larger than, the actual wavelength in the beam were used, the resulting displacement would be practically zero, and

thus eqn. (3.5 *a*) shows clearly how powerful the technique of interaction between travelling fields and electrons can be.

Similarly, it can be shown that electrons travelling in an axial field E_z will be velocity-modulated and bunched, resulting in a convection current component of amplitude

$$i = \frac{\omega}{u_0} \frac{eE_z z^2}{2mu_0^2} [(\sin\theta - \theta)^2 + (1 - \cos\theta)^2]^{1/2} \theta^{-2}, \qquad \ldots\ldots(3.7)$$

reducing to

$$i = \alpha(4\pi^2 z^2/\lambda^2)(V/4V_0) \qquad \ldots\ldots(3.7a)$$

when $u_0 = v$.

We can compare the velocity, or energy modulation, obtained with that of a conventional resonator or rhumbatron, and it can be shown that the helical transmission line is described by an equivalent shunt impedance

$$Z_{equ} = (4\pi^2 z^2/\lambda^2)\alpha^2 Z. \qquad \ldots\ldots(3.8)$$

With reasonable values, such as $\alpha = 0{\cdot}7$, $z = 20$ cm, $\lambda = 0{\cdot}7$ cm and $Z = 300$ ohm we have $Z_{equ} = 4{\cdot}8 \times 10^6$ ohm. This is many orders of magnitude higher than can be obtained with practical resonators at microwave frequencies, mainly on account of transit time and bandwidth requirements. Furthermore, it is apparent at once that Z_{equ} will change with frequency very much more slowly than the shunt impedance of a resonant cavity with a Q factor normally in the region between a hundred and a few thousand.

After calculating the effect of a wave on the beam a calculation was made of the influence of a displacement-modulated and of a density-modulated electron beam on the helical transmission line.

Let us describe a displacement-modulated beam, with an electron current I_0, by $s = s_0 \sin(\omega t - 2\pi z/\Lambda)$; Λ is the 'wavelength' of the electron beam travelling with a velocity u_0. It can be shown that voltage waves are induced in the transmission line both in the forward and backward directions. The forward voltage has an amplitude

$$V_2 = \tau \frac{\pi I_0 Z s_0 z}{D\Lambda} \frac{\sin \frac{1}{2}\theta}{\frac{1}{2}\theta}, \qquad \ldots\ldots(3.9)$$

where θ is given by (3.6).

Similarly it can be shown that an electron beam density-modulated uniformly along distance with a convection current of amplitude i and wavelength Λ induces a voltage in the transmission line of amplitude

$$V_2 = \alpha \frac{\pi i Z z}{\Lambda} \frac{\sin \frac{1}{2}\theta}{\frac{1}{2}\theta}. \qquad \ldots\ldots(3.10)$$

It will thus be seen that, when $u_0 = v$, beam modulations constant with distance give rise to induced voltage waves in the transmission line growing linearly with distance. On the other hand, we saw that a wave travelling along the transmission line with uniform amplitude gave rise to beam modulations growing with the square of distance.

It seemed therefore of interest to enquire what would happen if the density-modulation on a beam, due to a primary (uniform) voltage wave V on the transmission line, induced a wave on the transmission line. The amplitude of this wave, called the secondary wave, can, when $u_0 = v$, easily be shown to be

$$V_2 = \frac{1}{3!} \frac{\alpha^2 Z I_0}{4V_0} \left(\frac{2\pi z}{\lambda}\right)^3 V, \qquad \ldots\ldots(3.11)$$

and it can be shown that the phase of this wave lags 90° behind the primary wave V. The vector sum of a voltage V_2 at 90° to V and of V represents a total

voltage amplitude of $(V^2+V_2^2)^{1/2}$, which is larger than V. Thus it was realized that one of the results of the interaction between a wave and an electron beam might be a direct increase of the wave with distance.

In order to verify the basic assumptions on which the calculations were based, and to find the order of magnitude of some of the quantities such as the field weakening factors τ and α the following experiments were devised.

Radio-frequency power of approximately 10 cm wavelength was fed into a helix 21 cm long and propagating a predominantly travelling wave with a phase velocity $\frac{1}{10}\boldsymbol{c}$, using a construction which enabled a narrow beam of electrons to be shot either through the space between the helix and the outer conductor or along the axis of the helix (fig. 1). In the former case the beam was focused on

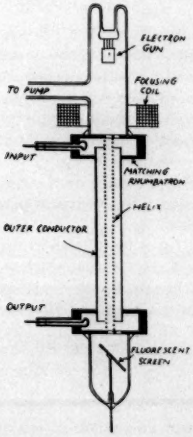


Fig. 1. Sketch of the experimental arrangement used for demonstrating the interaction between travelling waves and an electron beam.

to a fluorescent screen where it formed a luminous spot, in the latter case the beam could be made to traverse a resonant cavity, tuned to the signal frequency, or collected at an electrode, the potential of which could be varied.

In the first case, on increasing the beam voltage, the luminous spot drew out into a radial line, reaching a maximum amplitude at about 2 500 volts and showing a variation of amplitude in qualitative agreement with eqn. (3.5). From the known quantities of the system it was possible to obtain an average value of τ, which turned out to be in the region of 0·25.

In the second case the r.f. energy increment of the beam was measured by means of the biased collector, and again agreement was obtained with eqn. (3.5), leading to an effective value for α of 0·16.

The confirmation of the major basic assumptions of the theory encouraged an attempt to verify eqn. (3.11), which was done by measuring the r.f. power

which emerged from the output end of the 21 cm long helix as the beam current and voltage were varied. With the beam shot along the axis of the helix the result was an increase in the r.f. power from the output by 50% at a beam current of one milliampere and a beam voltage of 2440 volts, and a decrease of 40% when the beam voltage was 2200 volts.

When the beam was shot between the helix and the outer conductor the effects were very much smaller though similar.

On the whole these experiments confirmed the theory leading up to eqn. (3.11) and gave values of τ and α in agreement with the earlier experiments.

Using a microwave receiver whose sensitivity (expressed in terms of the noise factor) had been calibrated, the noise power due to shot noise in the electron beam emerging from the helix was measured and it was found that it was necessary to assume that the electron beam was less than completely random by a factor Γ^2 of about 0·1 to reconcile this measurement with the theory of the experiment.

The next step was, obviously, to try to make a proper amplifier relying on the theoretically predicted and now experimentally observed fact that the wave travelling along the helix grows on interacting with the beam.

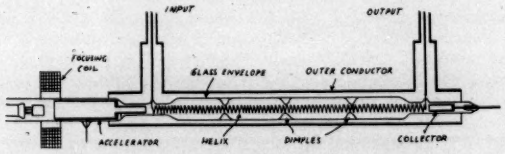


Fig. 2. Sketch of the first travelling-wave tube amplifier.

Theory showed that the simplest way of increasing the gain of the tube was to make it longer; eqn. (3.11) showed indeed that the r.f. power in the helix should grow with sixth power of the distance. It was realized that a theory taking only two terms of an infinite series could not be very reliable; nevertheless it can easily be shown that next higher order waves would become important only in the case of gain much higher than was envisaged at the time.

Similarly, theory showed that the noise factor of the tube should decrease with increasing length. Hence it was decided to make the tube as long as was physically possible under the conditions of restricted laboratory space, resulting in a length of helix of 66 cm.

Experiments on the attenuation of copper helices supported in close-fitting glass tubes showed that it was necessary to keep the glass away from the helix if the attenuation was to be less than, say, 10 db/metre. At this time the obvious —though, as it turned out later, false—view was held that attenuation is wholly bad and should be reduced as much as possible, and therefore a dimpled glass tube, as shown in fig. 2, was used as the support for the helix, to prevent it from sagging and vibrating. The 'cold' overall insertion loss of the tube was 3·5 decibels, due to losses in the copper and glass and mismatches at input and output.

With careful magnetic shielding an electron beam was projected along the axis of the 0·65 cm diameter helix, and 110 μA out of a total current of 130 μA reached the collector with the beam voltage in the neighbourhood of 2000 volts.

When the beam voltage was 1 830 volts, a net power gain of 6 was obtained, and the overall noise factor of tube plus microwave receiver was improved by 2 db from 16 db to 14 db above ideal.*

At higher beam voltages the tube generated oscillations near 3 300 Mc/s, the frequency used in these experiments; these oscillations sometimes persisted with beam currents as low as 20 μA.

A second tube, in which the helix of smaller diameter was supported in a dimpled fused-quartz tube (to reduce the insertion loss still further) gave a net power gain of 14 with 40 μA at 1 400 volts, and the noise factor of the tube by itself corresponded to 11 db above ideal, coming close to the sensitivity of a good crystal mixer–intermediate-frequency amplifier combination.

No particular attention was paid at this time to the bandwidth of the device once a quick experiment had shown that a variation of the signal frequency by

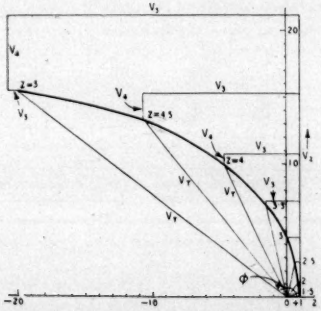


Fig. 3. Plot of the amplitude of the higher-order waves as a function of the parameter $x=(2\pi NC)$ in the complex plane. ϕ is the phase of the amplified wave with respect to the original wave of unit amplitude.

60 Mc/s, leaving all other parameters constant, had little effect on the gain of the tube. Since a bandwidth of 6 Mc/s is adequate for a radar amplifier, no further thought was given to bandwidth.

The theory, as developed so far, was clearly insufficient to describe the process of amplification as observed, with gains higher than 10, and was carried further by including all the higher-order waves. It will be recalled that a primary wave of amplitude constant with distance gives rise to a secondary wave increasing with the third power of distance, and it can also be shown that the secondary wave lags 90° behind the primary one. Similarly it can be shown that the secondary wave gives rise to a tertiary wave lagging 90° behind the secondary one, and growing with the sixth power of distance, and so forth. The complete series of the waves has been represented in graphical form in fig. 3 and can be written in the form

$$\Sigma V = V\left[1 + j\frac{(2\pi NC)^3}{3!} + j^2\frac{(2\pi NC)^6}{6!} + \ldots\right], \quad \ldots\ldots(3.12)$$

* See D. K. C. MacDonald, *Rep. Prog. Phys.*, 1948–9, **12**, 70, for the definition of noise factor.

where $N=z/\lambda$, the number of wavelengths along the tube, and $C=(\alpha^2 Z I_0/4V_0)^{1/3}$. This can be separated into a real and an imaginary part $\Sigma V=V[R+jX]$ and written in closed form

$$R=\tfrac{2}{3}\cosh\{(\sqrt{3}/2)(2\pi NC)\}\cos\{\tfrac{1}{2}(2\pi NC)\}+\tfrac{1}{3}\cos\{2\pi NC\}, \quad \ldots\ldots(3.13)$$

$$X=\tfrac{2}{3}\cosh\{(\sqrt{3}/2)(2\pi NC)\}\sin\{\tfrac{1}{2}(2\pi NC)\}-\tfrac{1}{3}\sin\{2\pi NC\}. \quad \ldots\ldots(3.14)$$

In order to calculate the power gain we need the square of the amplitude of ΣV, and this can be written, in closed form,

$$\{\Sigma V\}^2=\tfrac{1}{9}\{[2\cosh(\sqrt{3}/2)(2\pi NC)+\cos\tfrac{3}{2}(2\pi NC)]^2+\sin^2\tfrac{3}{2}(2\pi NC)\}V^2. \quad ..(3.15)$$

For values of $2\pi NC$ larger than 3 this tends to $\{\Sigma V\}^2=\frac{1}{9}\exp\{\sqrt{3}\,.\,2\pi NC\}V^2$. We can thus write for the power gain $\{\Sigma V\}^2/V^2=G$ in decibels,

$$G=10\log_{10}\{\tfrac{1}{9}[\exp(\sqrt{3}\,.\,2\pi NC)]\text{ decibels.} \quad \ldots\ldots(3.16)$$

For small values of $2\pi NC$ we use only the first two terms of the series for ΣV and find the initial voltage gain

$$G_{\text{initial}}=1+\tfrac{1}{12}(2\pi NC)^3. \quad \ldots\ldots(3.17)$$

The sum of all the waves ΣV, or the total wave, lags behind the initial wave by an angle ϕ given by $\tan\phi=X/R$, which tends to $\frac{1}{8}(2\pi NC)^3$ for small values of $2\pi NC$ and to $\tan(\frac{1}{2}\,.\,2\pi NC)$ for large values of $2\pi NC$. This means that the total wave, starting off with the same velocity v as the wave in the absence of the beam, is at first retarded at a rate proportional to N^2, later settling down to a constant velocity v' which is lower than v by a fraction $\frac{1}{2}v\,.\,2\pi NC$.

A similar calculation was carried out for the infinite series of waves induced in the helix by the shot noise current in the beam, resulting in a closed expression similar in character to eqn. (3.15). Since we are only interested in an amplifier of relatively high gain, when $2\pi NC$ is larger than 3, the total shot noise voltage can also be expressed sufficiently well by a simple exponential function, and on this basis one obtains a simple expression for the noise factor:

$$F=1+(2\boldsymbol{e}V_0/\boldsymbol{k}T)\Gamma^2 C, \quad \ldots\ldots(3.18)$$

where $\boldsymbol{e}$ is the electron charge, $1{\cdot}602\times10^{-19}$ coulomb, $\boldsymbol{k}$ is Boltzmann's constant, $1{\cdot}38\times10^{-23}$ joule/degree. Γ^2, as has been mentioned before, is the so-called space-charge smoothing factor, the fraction by which the shot fluctuation in the electron beam is less than completely random.

Taking for T the standard reference temperature 290° K, we can write

$$F=1+80\Gamma^2 V_0 C.$$

It is of interest to see how these expressions compare with the experimental results obtained so far. Suppose we assume the following values: $\alpha=0{\cdot}33$, $Z=200$ ohm, $V_0=1\,400$ volts, $I_0=4\times10^{-5}$ ampere. This gives $C=(\alpha^2 Z I_0/4V_0)^{1/3}=5{\cdot}4\times10^{-3}$. The number of wavelengths N was about 90; thus we have for $2\pi NC$ approximately 3, and the net gain $G=13{\cdot}1$ db. The observed gain was 14, that is 11·5 db. The noise factor depends very much on the value of Γ^2 obtaining in the experiment, and thus it may be more meaningful to assume the correctness of the theory and calculate Γ^2 instead, given the experimental noise factor. This procedure gives, with a measured noise factor $F=12{\cdot}6$, $\Gamma^2=(F-1)/80V_0C\simeq0{\cdot}02$. Thus it will be seen that again the electron beam appears to be much less random than a completely random electron stream,

that is to say, one in which the mean square current fluctuation per unit bandwidth is given by $\overline{i^2} = 2eI_0$.

On the whole it therefore seemed that the theory could give a satisfactory account of the experiments. Least certain was the field distribution inside the helix, and measurements were therefore made with a helix scaled up linearly by a factor 10. The axial field was explored with a small probe and crystal detector and good agreement was found with a theory given by Dr. H. Motz (at that time at the Engineering Laboratory, Oxford) in an unpublished note. It was established that the axial field at a distance r from the axis can be described by an expression of the form $E_z = E_{za} I_0(\gamma r)/I_0(\gamma a)$, where E_{za} is the axial field very close to the helix of radius a and I_0 is the modified Bessel function. γ is a propagation constant and very nearly equal to $2\pi/\lambda$, where λ is the wavelength along the helix. Thus the field weakening factor α for a thin beam on the axis of the helix could be identified with $[I_0(\gamma a)]^{-1}$.

§ 4. DESCRIPTION OF MECHANISM OF TRAVELLING-WAVE TUBE GAIN

At this stage it may well be asked: what is the physical mechanism which causes a wave interacting with an electron beam at the same speed to increase? How is the transfer from the kinetic energy of the beam to the radio-frequency energy of the wave accomplished?

In the case of the klystron these questions are easily answered: The r.f. input power into the buncher causes an r.f. field to appear across the buncher gap which velocity-modulates the beam. In a first approximation no energy is needed to velocity-modulate a beam, since as many electrons are accelerated as retarded. The velocity modulation is turned into a current amplitude modulation in the drift-tube.* As this a.c. current passes through the catcher it induces a current therein which in turn produces an a.c. field across the catcher-gap in such a phase that, when more electrons pass through it, it is retarding and when less electrons pass through it, it is accelerating. Therefore more electrons are retarded than accelerated, and on the average the electron beam loses energy. This energy appears in the form of radio-frequency energy in the catcher, and part of it becomes available as output power, and may exceed the input power necessary to produce the initial r.f. field across the buncher gap. Thus the klystron amplifies.

It can be shown that the process of amplification in the travelling-wave tube can be explained on very similar lines.

The r.f. input power travels along the helix in the form of an r.f. field; an electron beam travelling in this field with the same velocity will, in the first place, experience forces causing it to become velocity-modulated. As in the klystron, no energy is transferred from the wave to the beam in this process, since as many electrons are accelerated as retarded. This velocity modulation is at once turned into a current-amplitude modulation, growing with the square of distance, and lagging 90° behind the wave.

This current modulation induces, at once, a secondary wave in the helix, its amplitude growing with the cube of distance, as has been shown. The important point to note, however, is that the current modulation induces a secondary wave

* It is worth noting that, if space-charge forces are neglected, there is no r.f. energy in a bunched electron beam.

in such a phase that, where there is an excess of electrons over the mean, the phase of the wave is retarding, and where there is a deficit the wave is accelerating. Thus more electrons are retarded than accelerated, and on the average the electron beam loses energy. This energy is fed into the wave, and thus the travelling-wave tube amplifies.

This process does not, of course, stop with the secondary wave. The secondary wave causes a secondary current modulation in the beam which gives rise to a tertiary wave, and so on. At every stage, however, no energy is taken from the wave in the process of velocity-modulating and bunching the beam, while energy is fed into the helix from the beam in the process of inducing a wave in the helix. Viewing all the component waves as one complete wave—the resultant of the interaction between the initial wave and the beam—the complete wave will be slower than the initial one, since it is made up of components each of which lags 90° on its predecessor. Thus the seeming paradox of a net energy transfer between a wave and electrons travelling at the same speed is resolved—the real wave is slower than the beam.

It will be appreciated that these physically intuitive descriptions are only approximately true, and fail to take into account a number of factors such as the finite amplitude of signals, space-charge forces and circuit losses. Taking these into account would, however, so complicate matters as to obscure the physical processes completely, and they have therefore been neglected.

§ 5. INTERIM REVIEW OF THE PROBLEM

With these investigations, the fundamental fact that interaction between an electron beam and a wave travelling at the same velocity leads to an increase of the wave, approximately exponential with distance, had been established.

It had been shown, in principle, that sufficient gain, sensitivity and bandwidth could be obtained with the travelling-wave tube (as it came to be called) for it to be eventually of use as an r.f. amplifier in connection with radar receivers. Nevertheless, it has to be realized that the work described so far was done in 1943 and 1944; to develop a new type of vacuum tube to the point where it can be used in practice needs a great amount of scientific and technological effort, stretching over years in some cases. The development of the multi-cavity magnetron (Boot and Randall 1946) can be cited as an example, and the magnetron was the very heart of centrimetric radar. The travelling-wave tube did not at the start promise to do anything that could not be done already as well, or better, with existing means, and therefore the effort that was made available in war-time towards its development was very limited. Thus progress was very slow once the principle had been established, and it was only towards the end of the war, when J. R. Pierce, of Bell Telephone Laboratories, in the United States, realizing the potential importance of such a tube for microwave communication circuits, took an interest in the travelling-wave tube, that considerable and important progress was made.

The chief problems awaiting immediate attention were:

What is the influence of attenuation in the helix?

What is the effect of a beam velocity differing from the undisturbed wave velocity?

What will be the effect of large signals on the interaction between beam and wave; in particular, what is the highest power output efficiency one can expect?

What is the effect of space charge?

How can oscillations, due to feedback from the output via the helix to the input, be prevented?

How can the sensitivity of the tube be improved; that is to say, what is the mechanism of space-charge smoothing and partition noise at microwave frequencies?

What is the bandwidth over which amplification can be obtained with a single tube?

Important long-term objectives also became apparent as time went on. These were:

Investigations into other types of travelling-wave electron stream interactions.

Investigations into structures other than the helix capable of supporting slow waves.

Investigations into the range of frequencies over which the principle of travelling-wave interaction can be applied with success.

By now some sort of answers to all these questions have been obtained. Some of the problems have given rise to a tremendous amount of theoretical work, leading to similar results in many cases, and in the following only representative treatments will be illustrated. The amount of published experimental work is not quite so large but is continuously being overtaken by later developments, and this account may well be partially out of date by the time of its appearance. Again, here we shall attempt to report mainly on representative results, corresponding to definite steps in the development. In a field that has obviously not reached its furthest limits in many respects, this is the only possible course. It is impossible, at this stage, to assess the degree of importance which is to be attributed to the various discoveries, devices and developments, and to put them in their proper perspective. The best that can be done is to point out possible applications to astronomy or physics, engineering or technology, as the case may be, and as the occasion arises.

§6. DEVELOPMENTS AT THE BELL TELEPHONE LABORATORIES

In the summer of 1946 J. R. Pierce announced the results of investigations carried out at Bell Telephone Laboratories in the United States. It appeared that the work there had led to the development of travelling-wave tubes with the following typical characteristics: beam voltage, 1 670 volts; cathode current, 8 mA; gain at 3 600 Mc/s, 23 db; bandwidth between 3 db, 800 Mc/s; maximum power output, 200 milliwatts; transmission loss with beam off, 33 db.

A sketch of such a tube with its associated circuits is shown in fig. 4. Magnetic fields are used to prevent spreading of the electron beam. The helix is held between four ceramic rods and is about 11 inches long. Matching to the waveguides is achieved by attaching the helix to short stubs, which project across the waveguides parallel to the electric field there. The stubs, in turn, are attached to collars which fit snugly into the glass envelope and form with it and the corresponding collars outside the envelope by-pass condensers for the r.f. currents, so that in effect the helix is attached to the waveguide wall through the short stub. The remarkable feature of this tube is, of course, its extremely wide bandwidth in conjunction with its high gain.

There had been in existence what amounted to a theorem on the product of gain times bandwidth obtainable with conventional electron tubes (Wheeler 1939, Hansen 1945). Here was a tube that exceeded the limit set by the theorem

by about two orders of magnitude. The wide bandwidth had been achieved by proper attention to the match between the helix and input and output waveguides, and it appeared that such junctions could be made of very wide bandwidth indeed. Even so, these junctions set the limit to the bandwidth, the system helix–electron beam having inherently a still wider bandwidth. The gain will fall off at higher frequencies because the fields inside the helix become weaker near the axis. At lower frequencies the tube will be shorter in terms of wavelengths.

The large gain (as compared with the author's early results) is obtained by the ingenious use of attenuation along the helix. In the early tubes an increase in beam current was normally accompanied by an increase in gain up to the point where the tube started to oscillate, and this point was often reached with much lower beam currents than could actually be produced. This limited the gain severely. It was an important discovery, a consequence of Pierce's extension of the theory to include loss in the helix, that such loss had comparatively little

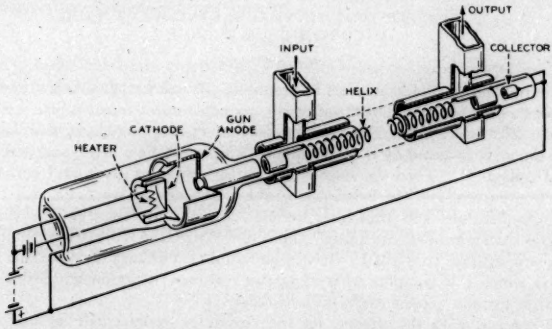


Fig. 4. Travelling-wave tube with its associated circuits.
Acknowledgment to Bell Telephone Laboratories.

effect on the 'forward gain' of a travelling-wave tube. A backward travelling wave, say, due to reflection at the output junction, may thus be attenuated more than a forward travelling wave will be amplified, and thus stable operation is possible with forward gains of the same order of magnitude as the backward loss—up to a limit.

It was clear that the enormous bandwidth demonstrated by Pierce and collaborators opened up possibilities in the field of communications never before envisaged, and the immediate result of the announcement was attempts in a number of places to reproduce the results announced by Pierce. Another result was a great number of attempts in various places to derive a theoretical description of the process of amplification in the travelling-wave tube, a task inviting to mathematically minded physicists. The latter attempts were far more uniformly successful than the former, and very soon after Pierce's announcement there appeared a number of papers (Pierce 1947, Chu and Jackson 1948, Bernier 1947 a, b, Rydbeck 1948, Shulman and Heagy 1947, Blanc-Lapierre, Lapostolle and Wallauschek 1947, Goudet 1948) containing treatments leading to solutions

identical with the ones given earlier by the author. The methods of attack, however, were in general different and far more elegant, and fall roughly into two classes: one class might be called a 'network and electron beam' theory, and is exemplified by Pierce's and Bernier's theories; the other might be called 'waveguide and electron stream' theory, and is exemplified by Chu and Jackson's treatment. The former class is more useful to the designer and user of travelling-wave tubes since it leads to a number of simple equations in which the significance of all relevant parameters is easily understood. The latter class is, in principle, far more elegant and rigorous, but does not lead to such simple and explicit relations as the former, although the general results of both classes are in fair agreement, qualitatively and quantitatively. This fact, however, gives one confidence in the 'network and electron beam' type of theory, and below we shall sketch an outline of such a theory based on that given by J. R. Pierce (1947) and amplified in his book (1950).

§ 7. THEORY OF THE TRAVELLING-WAVE TUBE ACCORDING TO PIERCE

7.1. *Basic Equations*

The method of attack consists in assuming that all a.c. quantities vary with distance as $\exp(j\omega t - \Gamma z)$, where Γ is the propagation constant which we want to find. First we assume a convection current i along the beam, and we shall derive the wave induced by it in the circuit (which may be a helix) and described by the voltage V. Then we assume the voltage wave as given and derive the convection current i in the beam due to the action of this wave, i.e. due to the velocity-modulation and bunching action. Finally, we identify the derived convection current with the initial convection current and obtain an equation for the propagation constant Γ. It will be found that among the solutions there is one in which Γ is complex with a negative real part, which means that a wave exists which grows exponentially with distance.

To proceed with the theory, let the circuit be represented by uniformly distributed series inductance X and shunt susceptance B, with current I and voltage V. The electron beam is close to the circuit and has a convection current i flowing in it. This will impress a displacement current $J = -\partial i/\partial z$ on the circuit and we have thus the following equations for the network:

$$\frac{\partial I}{\partial z} = -jBV + J; \qquad \frac{\partial V}{\partial z} = -jXI. \qquad \ldots\ldots(7.1)$$

Recalling that all a.c. quantities vary with distance as $\exp(-\Gamma z)$, we can write

$$-\Gamma I = -jBV + \Gamma i; \qquad -\Gamma V = -jXI. \qquad \ldots\ldots(7.1a)$$

Introducing $\Gamma_1 = j(BX)^{1/2}$, the propagation constant of the circuit, and $K = (X/B)^{1/2}$, the characteristic impedance of the circuit, in the absence of electrons, and eliminating I we obtain

$$V = -\Gamma\Gamma_1 Ki/(\Gamma^2 - \Gamma_1^2). \qquad \ldots\ldots(7.2)$$

This is the voltage induced in the circuit by the convection current i in the beam.

To find the effect of the voltage in the circuit on the beam we proceed as follows: According to Newton's law the force on an electron is $\boldsymbol{m}d(u_0+v)/dt = \boldsymbol{e}E_z$, where u_0 is the mean velocity of the electron, v the a.c. component of velocity, and E_z the axial electric field ($=\partial V/\partial z$).

The change in the a.c. component of velocity can be expressed in terms of partial derivatives

$$\frac{dv}{dt} = \frac{\partial v}{\partial t_s} + \frac{\partial v}{\partial z_t}\frac{dz}{dt}. \qquad \ldots\ldots(7.3)$$

But $dz/dt = u_0 + v$, and thus we can write

$$\frac{\partial v}{\partial t} + \frac{\partial v}{\partial z}(u_0 + v) = \frac{e}{m}\frac{\partial V}{\partial z}. \qquad \ldots\ldots(7.4)$$

In a small-signal theory v can be neglected in comparison with the mean velocity u_0, and carrying out the differentiations we write, therefore, $(j\omega - u_0\Gamma)v = -(e/m)\Gamma V$ or $v = -e\Gamma V/mu_0(j\beta_e - \Gamma)$, where $\beta_e = \omega/u_0$, the electron beam propagation constant.

The total convection current in the beam is given by

$$-I_0 + i = (u_0 + v)(\rho_0 + \rho) \qquad \ldots\ldots(7.5)$$

and, neglecting products of small quantities,

$$i = \rho_0 v + u_0\rho, \qquad \ldots\ldots(7.5a)$$

where I_0 is the d.c. beam current, ρ_0 is the d.c. beam charge density ($= -I_0/u_0$), ρ is the a.c. beam charge density.

To find ρ we make use of the equation of conservation of charge, which, in a one-dimensional system, is given by $\partial i/\partial z = -\partial\rho/\partial t$, $-\Gamma i = -j\omega\rho$; hence

$$i = \frac{jI_0\beta_e\Gamma V}{2V_0(j\beta_e - \Gamma)^2}, \qquad \ldots\ldots(7.6)$$

remembering that $\frac{1}{2}mu_0^2 = eV_0$, where V_0 is the mean beam voltage. Equation (7.6) is the convection current in the beam due to a voltage V in the circuit.

If i in (7.6) is the same as in (7.2), i.e. if the scheme is consistent, Γ must satisfy the equation

$$1 = \frac{jKI_0\beta_e\Gamma^2\Gamma_1}{2V_0(\Gamma_1^2 - \Gamma^2)(j\beta_e - \Gamma)^2}. \qquad \ldots\ldots(7.7)$$

This equation is of the fourth degree, and we therefore expect four values of Γ, that is to say, four waves.

7.2. *Propagation neglecting Space Charge and Circuit Attenuation; Synchronous Velocities*

Assuming, to start with, that the wave in the absence of electrons has the same speed as the electrons, we put $-\Gamma_1 = -j\beta_e$. This also implies that there is no attenuation in the circuit. The situation when the velocity is different from the beam velocity, and when there is a finite attenuation, will be dealt with later, in order not to obscure the argument at this point. Now let us define a new parameter

$$C = (KI_0/4V_0)^{1/3}, \qquad \ldots\ldots(7.8)$$

and let us describe a forward travelling wave by $-\Gamma = -\Gamma_2(1 + jC\delta)$. Normally C will be so small compared with unity that the product $C\delta$ is also small compared with unity; δ, which is now the parameter we want to find, need not be small.

Equation (7.7) can now be written

$$1 = \frac{2C^3(1 + jC\delta)^2}{(1 + \frac{1}{2}jC\delta)(-jC\delta)^3}, \qquad \ldots\ldots(7.7a)$$

and with the approximation $C\delta \ll 1$ we find $C\delta = (-j)^{1/3}$. This has three roots, δ_1, δ_2 and δ_3, representing three forward waves:

$$\delta_1 = \tfrac{1}{2}(\sqrt{3}-j), \qquad \delta_2 = \tfrac{1}{2}(-\sqrt{3}-j), \qquad \delta_3 = j. \qquad \ldots\ldots(7.9)$$

The fourth root has been eliminated by the approximations made. Let us put for a backward wave $-\Gamma = +\Gamma_1(1+C\delta)$. This yields $\delta_4 = -\frac{1}{4}C^2$, which is negligible compared with unity and with the other values of δ. Thus the backward wave is practically the same as in the absence of electrons, which was to be expected.

Thus the three forward waves are described by

$$\exp(j\omega t - \Gamma z) = \exp(j\omega t - j\beta_e z)\exp(\delta C\beta_e z). \qquad \ldots\ldots(7.10)$$

We see that the first wave (using δ_1), is an increasing wave which travels a little more slowly than the electrons. The second wave is a decreasing wave which also travels a little more slowly than the electrons. The third wave is an unattenuated wave which travels faster than the electrons.

In order to find the actual gain of an applied signal we have to discuss the boundary conditions. In our problem we have at the beginning of the helix a circuit voltage V which must be the same in the presence or absence of an electron beam. Both the convection current i and the a.c. velocity v are normally assumed to be zero. The convection current i is given by (7.6) which, on substituting for Γ, becomes $i = -I_0 V/2V_0 C^2\delta^2$. Similarly, we can express the velocity in terms of V and δ: $v = (\boldsymbol{e}/\boldsymbol{m})V/ju_0 C\delta$.

The circuit voltage V must be made up of the three forward waves V_1, V_2 and V_3, characterized by the three values of δ. Thus we must have $V_1 + V_2 + V_3 = V$. The convection current and a.c. velocity must also be split up into three components: $i_1 + i_2 + i_3 = 0$, $v_1 + v_2 + v_3 = 0$. Substituting for i and v we have

$$\frac{V_1}{\delta_1^2} + \frac{V_2}{\delta_2^2} + \frac{V_3}{\delta_3^2} = -\frac{2V_0C^2}{I_0}i = 0 \quad \text{and} \quad \frac{V_1}{\delta_1} + \frac{V_2}{\delta_2} + \frac{V_3}{\delta_3} = j\frac{u_0 C}{\boldsymbol{e}/\boldsymbol{m}}v = 0.$$

Solving for V_1 we find $V_1 = V[(1-\delta_2/\delta_1)(1-\delta_3/\delta_1)]^{-1}$; $V_1 = \frac{1}{3}V$. Similarly we find $V_2 = V_3 = V_1 = \frac{1}{3}V$. V has been used to denote the voltage at $z = 0$. If V_z is the voltage at z we have

$$V_z = \tfrac{1}{3}V\exp[-j\beta_e(1-C)z]\{1 + 2\cosh[(\sqrt{3}/2)\beta_e Cz]\exp[-j(3/2)\beta_e Cz]\}, \qquad \ldots\ldots(7.11)$$

which will be seen is identical with expressions (3.13) and (3.14).

The power gain is, of course, identical with expression (3.15) given earlier. After a certain distance of interaction between waves and beam, the increasing wave predominates, and one can write the power gain, in decibels,

$$G = A + BCN, \qquad \ldots\ldots(7.12)$$

where $A = -9{\cdot}54$ db, $B = 47{\cdot}3$ and $N =$ number of cycles in the tubes. A plot of the power gain is given in fig. 5, and we see that the asymptotic expression (7.12) is a good approximation for CN greater than 0·4.

A similar analysis, starting off with an initial a.c. convection current equal to the shot-current in the beam, leads to the same expression for the noise factor as given earlier (3.18), and need therefore not be repeated. The real importance of the method outlined above lies in its ability to describe the process of amplification when we do *not* neglect the effects of space charge, loss in the helix or of phase velocity different from the electron velocity. These are now obtained

by the fundamentally simple device of making slight modifications in the propagation constant; with the author's earlier method of describing the interaction by an infinite series of higher-order waves, the complexity of the problem defies analysis.

7.3. *Propagation taking Space Charge and Circuit Attenuation into Account; Non-Synchronous Velocities*

The effect of space charge in the beam manifests itself by the appearance of a.c. fields around the bunched beam even if there is no wave on the circuit. This can be represented in our picture of a beam travelling close to a transmission line (with distributed series inductance X and shunt susceptance B) by a distributed capacity C_1 between the beam and the circuit. The convection current i in the

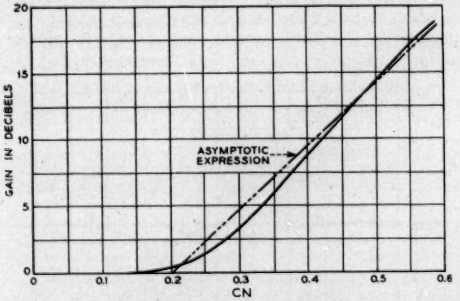


Fig. 5. Power gain in decibels as function of CN, as given by eqn. (3.15) and by the asymptotic expression (7.12).

Acknowledgment to Bell Telephone Laboratories.

beam and the a.c. charge in the beam are connected by $i = j\omega\rho/\Gamma$ by the principle of conservation of charge. Therefore the voltage across the distributed capacity C_1 is $V_c = \rho/C_1 = \Gamma i/j\omega C_1$, and the voltage on the circuit due to an impressed current i in the beam has now to be written

$$V = \left[\frac{\Gamma\Gamma_1 K}{\Gamma_1^2 - \Gamma^2} - j\frac{\Gamma}{\omega C_1}\right] i.$$

Together with (7.6) the propagation constant is determined, as before. Now, however, we shall be also interested in values of Γ_1, which differ from β_e by small amounts, and we write accordingly

$$-\Gamma = -j\beta_e + \beta_e C\delta, \qquad -\Gamma_1 = -j\beta_e - j\beta_e Cb - \beta_e Cd.$$

Γ describes the propagation in the presence of electrons, and a positive real part of δ means an increasing wave, a positive imaginary part means a wave faster than the electrons.

Γ_1 describes the propagation in the absence of electrons, and a positive value of b means that the undisturbed wave travels *more slowly* than the electrons, while a positive value of d means an attenuated wave.

Solving the combined equations for V and i, after making the substitutions for Γ and Γ_1, and neglecting all terms multiplied with C as compared with unity, there results

$$\delta^2 = \frac{1}{(-b+jd+j\delta)} - 4QC, \qquad \ldots\ldots(7.13)$$

where $Q = \beta_e/2\omega C_1 K$ is the so-called space-charge parameter.

For simplicity we first assume $d = Q = 0$. Then $\delta^2(\delta + jb) = -j$, or, writing $\delta = x + jy$,

$$(x^2 - y^2)(y + b) + 2x^2 y + 1 = 0, \qquad x(x^2 - 3y^2 - 2yb) = 0.$$

Of the two roots of the second equation $x = 0$ corresponds to an unattenuated wave and $x^2 = 3y^2 - 2yb$ leads to an increasing and to a decreasing wave. Substituting for x^2 into the first expression gives $2yb^2 + 8y^2 b + 8y^3 + 1 = 0$, which is quadratic in b and can be solved by assigning various values of y. The values of y corresponding to $x = 0$ have been designated y_3; y_1 and y_2 correspond to the increasing wave x_1 and the decreasing wave x_2. Curves of x and y versus b are shown in fig. 6, which contains then the information needed to obtain all the

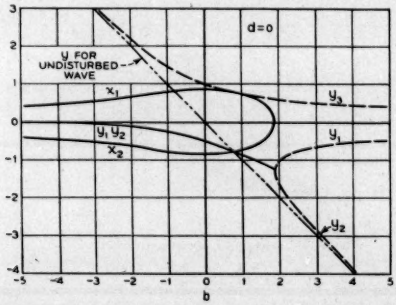


Fig. 6. Plot of the real and imaginary components x and y of δ as a function of b, a parameter proportional to the difference between electron and wave velocities.

Acknowledgment to Bell Telephone Laboratories.

three forward waves when the electron velocity differs from the phase velocity of the undisturbed wave. In particular, the curve labelled x_1 is directly proportional to the gain of the increasing wave in decibels, and it will be noted that maximum gain is obtained when $b = 0$; further, when $b > \frac{3}{2}2^{1/3}$ there is no increasing (or decreasing) wave. We recall that b is defined by $u_0/v_1 = 1 + Cb$, where u_0 is the electron velocity and v_1 is the phase velocity in the circuit in the absence of electrons.

A finite value of attenuation along the helix is defined by saying that, in the absence of electrons, the voltage decays with distance as $\exp(-\beta_e Cd)$. Hence the loss L in decibels per wavelength is $L = (20 \times 2\pi \log_{10} e)Cd$ or $d = 0{\cdot}01836\ L/C$.

One can now proceed as before and obtain plots of x and y as functions of b for various values of d. Such plots are given by Pierce (1947) and are instructive in showing how the waves are modified by the introduction of attenuation. It

appears that maximum gain still occurs when $b=0$, i.e. when electrons and wave travel with the same speed, but that its value is decreased as attenuation is increased. Indeed, it can be shown that, for small values of d, $x_1=x_{10}-\frac{1}{3}d$, where x_{10} is the value of x_1 for $d=0$.

We observe the important result that for small losses the *reduction of gain of the increasing wave* in decibels from the hypothetical gain with zero loss is one-third of the circuit attenuation, also in decibels. The reduction in *net gain* will be somewhat greater since the energy is now distributed differently between the three forward waves; this is taken care of by the term A in eqn. (7.12). We can write, when $b=Q=0$, approximately $A=-9{\cdot}6(1+\frac{1}{2}d)$ decibels.

This, then, is only a small correction; the reduction of the gain by one-third of the circuit loss, though not negligible, is not serious, as will be seen from the following consideration: suppose the circuit loss without beam is 45 decibels, and the gain of the forward wave without loss was 60 decibels. Then the gain of the forward wave will be $60-\frac{1}{3}.45=45$ decibels and the net gain about 10 decibels less still, i.e. 35 decibels. A tube with 35 decibels forward gain and 45 decibels backward loss will be stable under any conditions, and reflections due to mismatches at the ends of the tube cannot lead to oscillations.

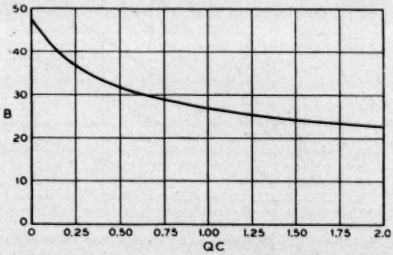


Fig. 7. Gain factor B as function of QC, the space-charge parameter.

Acknowledgment to Bell Telephone Laboratories.

To evaluate the effect of space charge we assume that there is no loss in the helix ($d=0$) and proceed, as before, with eqn. (7.13). This leads to cubic equations in x and y, and solutions have been found, as before, by numerical methods. Plots of the sets of x and y for the three forward waves for different values of QC are given by Pierce (1947), and it appears that as the space-charge parameter increases the value of b for maximum gain moves away from zero towards positive values of b. The maximum value of x_1, that is the gain in decibels of the forward wave itself, decreases somewhat as QC increases, as shown in fig. 7.

Q is the ratio of the capacitive impedance between the beam and the circuit to K, the circuit impedance, and has been plotted in fig. 8. There it is also shown that the product BQC (where B is the gain parameter of eqn. (7.12), which is proportional to gain per wavelength in decibels, varies linearly with QC when $QC<0{\cdot}025$ and with the three-fourths power of QC when $QC>0{\cdot}4$. This agrees well with results obtained by the 'waveguide and electron-beam' theories by more direct methods.

7.4. *Power Output and Efficiency*

We can arrive at a rough estimate of the maximum power output and efficiency if we assume that the maximum convection current in the beam i_{max} equals the beam current itself. We have

$$|i_{max}| = \frac{I_0}{2V_0C^2}\frac{|V_{max}|}{|\delta^2|}, \qquad |\delta^2| = 1.$$

Hence the maximum voltage in the circuit is $|V_{max}| = 2V_0C^2$ and the power in the circuit is $P_{out} = |V_{max}^2|/2K = \frac{1}{2}V_0I_0C$ since $C^3 = V_0I_0/4K$.

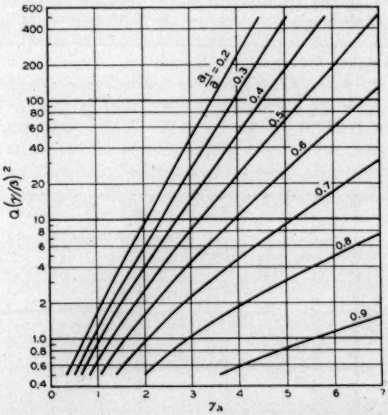


Fig. 8. Plot of $Q(\gamma/\beta)^2$ against γa for a hollow beam with radius a_1.

Acknowledgment to Bell Telephone Laboratories.

Thus the efficiency is of the order of $\frac{1}{2}C$. Subsequent work by Nordsieck (referred to by Pierce (1950, chapter on power output), by Pierce himself (1950) and by Slater (1950) show that efficiencies of $2C$ and more should be obtained, the efficiency increasing with the factor QC and as the beam velocity for optimum gain exceeds the undisturbed wave velocity.

We are here, of course, only concerned with the so-called 'electronic' efficiency, that is to say, the absence of loss in the circuit. Any such loss will reduce the overall efficiency, and can be calculated separately (Doehler and Kleen 1949 b).

7.5. *The Helix*

For a complete understanding of the travelling-wave tube of the helix type it is desirable to know some of the properties of the helix as a waveguide. In particular, one needs to know the nature and distribution of the fields inside and outside the helix, the phase velocity of waves propagated by the helix, and the impedance of the helix or, more specifically, what is the axial electric field strength associated with a given power flow. All this is dealt with by Pierce (1947) on the basic assumption that one can replace a wire helix by an infinitely thin

cylindrical sheet with infinite conductivity in a helical direction and zero conductivity at right angles to this, as shown in fig. 9. A multiple helix, that is, many parallel wires per turn, might be the closest approach to this model in practice. Rather surprisingly, however, the usual kind of wire helix, i.e. with one wire per turn, conforms fairly closely to the theory of the helical-sheet model.

The problem of the propagation of a wave along a helix has been treated in the past (Pocklington 1897, Sommerfeld 1907, Lenz 1914, Ollendorf 1925). Here we shall follow Pierce's treatment, which he ascribes to Schelkunoff.

An appropriate solution of the wave equation in cylindrical coordinates for a wave propagating in the z-direction with a velocity v less than the velocity of light, and having cylindrical symmetry, is

$$E_z = [AI_0(\gamma r) + BK_0(\gamma r)] \exp j(\omega t - \beta z),$$

where I_0 and K_0 are the modified Bessel functions and $\gamma^2 = \beta^2 - \beta_0^2 = (\omega/v)^2 - (\omega/\mathbf{c})^2$.

Fig. 9. Sketch of helically conducting sheet of radius a, conducting along a direction making an angle ψ with a plane normal to the axis.

Acknowledgment to Bell Telephone Laboratories.

Inside the helix of radius a the longitudinal components of the electromagnetic field must remain finite everywhere, and we therefore use only the I_0 and I_1 functions. Outside the helix the fields must vanish at infinity; thus, apart from proportionality factors, the fields will be described by the modified Bessel functions K_0 and K_1 of the second kind.

Expressing all quantities in terms of E_{za}, the longitudinal field at $r = a$, i.e. in the hypothetical helix surface, the individual field components inside the helix are

$$E_z = \frac{I_0(\gamma r)}{I_0(\gamma a)}, \qquad H_z = -j\frac{1}{120\pi}\frac{1}{\cot\psi}\frac{I_1(\gamma r)}{I_1(\gamma a)},$$

$$E_r = j\frac{\beta}{\gamma}\frac{I_1(\gamma r)}{I_0(\gamma a)}, \qquad H_r = \frac{1}{120\pi}\frac{1}{\cot\psi}\frac{\beta}{\beta_0}\frac{I_1(\gamma r)}{I_1(\gamma a)},$$

$$E_\varphi = -\frac{1}{\cot\psi}\frac{I_1(\gamma r)}{I_1(\gamma a)}, \qquad H_\varphi = j\frac{1}{120\pi}\frac{\beta_0}{\gamma}\frac{I_1(\gamma r)}{I_0(\gamma a)},$$

and outside the helix

$$E_z = \frac{K_0(\gamma r)}{K_0(\gamma a)}, \qquad H_z = j\frac{1}{120\pi}\frac{1}{\cot\psi}\frac{\gamma}{\beta_0}\frac{K_0(\gamma r)}{K_1(\gamma a)},$$

$$E_r = -j\frac{\beta}{\gamma}\frac{K_1(\gamma r)}{K_0(\gamma a)}, \qquad H_r = \frac{1}{120\pi}\frac{1}{\cot\psi}\frac{K_1(\gamma r)}{K_1(\gamma a)},$$

$$E_\varphi = -\frac{1}{\cot\psi}\frac{K_1(\gamma r)}{K_1(\gamma a)}, \qquad H_\varphi = -j\frac{1}{120\pi}\frac{\beta_0}{\gamma}\frac{K_1(\gamma r)}{K_0(\gamma a)},$$

all components propagating as $\exp j(\omega t - \beta z)$.

These quantities and the phase velocity are determined by the following boundary conditions: the tangential electric field must be perpendicular to the helix direction and continuous across the cylinder, and the tangential component

of magnetic field parallel to the helix direction must be continuous across the cylinder, since there can be no current perpendicular to this direction. The radial propagation constant γ can be obtained from the equation

$$\gamma a \left(\frac{I_0(\gamma a)K_0(\gamma a)}{I_1(\gamma a)K_1(\gamma a)}\right)^{1/2} = \beta_0 a \cot\psi. \qquad \ldots\ldots (7.14)$$

The expression within brackets, which we denote by $H(\gamma a) = (\beta_0/\gamma)\cot\psi$, is plotted in fig. 10.

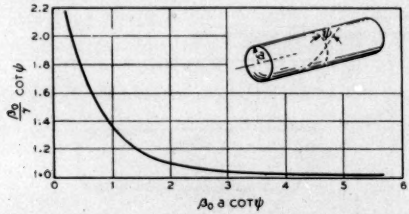


Fig. 10. $H(\gamma a) + (\beta_0/\gamma)\cot\psi$ as function of $\beta_0 a \cot\psi$.

Acknowledgment to Bell Telephone Laboratories.

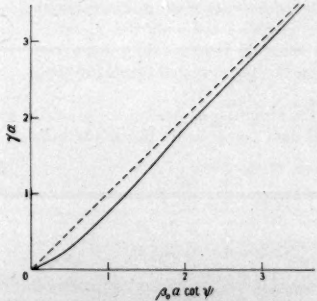


Fig. 11. γa plotted against $\beta_0 a \cot\psi$. γ is the radial propagation constant.

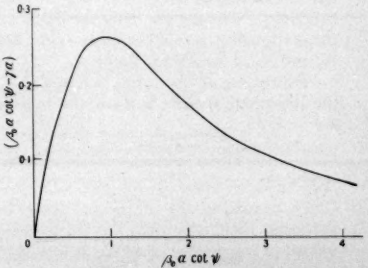


Fig. 12. $(\beta_0 a \cot\psi - \gamma a)$ plotted against $\beta_0 a \cot\psi$. This plot is useful in estimating dispersion of the wave.

When the phase velocity in the helix is small compared with **c**, γ/β is very nearly equal to unity, and eqn. (7.14) can be used to find the phase velocity directly. When $\beta_0 a \cot\psi$ is large compared with unity, the wave propagates with an axial velocity corresponding to a wave propagating along the helical direction with **c**. When $\beta_0 a \cot\psi$ is of the order of unity and less, the wave becomes progressively faster, and in this region the helix is said to be *dispersive*.

$\beta_0 a \cot\psi$ contains all the geometric parameters of the helix, and its variation with γa is shown in fig. 11; it is seen that γa approaches $\beta_0 a \cot\psi$ asymptotically. For greater accuracy, particularly with the smaller values of $\beta_0 a \cot\psi$, $\beta_0 a \cot\psi - \gamma a$ is plotted against $\beta_0 a \cot\psi$ in fig. 12.

The power propagated along a helix is given by

$$P = \tfrac{1}{2}\mathscr{R}\int E \times H^* \, d\tau$$

taken over a plane normal to the axis of propagation. When the appropriate terms for the radial and tangential fields are inserted one obtains a very complicated expression which will not be reproduced here. The quantity normally required when calculating the performance of a travelling-wave tube is the cube root of the effective impedance. Unfortunately the expressions and graphs given by Pierce refer to a parameter $E_z^2/\beta^2 P$ which is exactly twice K, which we have, rather loosely, called the impedance. To avoid misunderstandings, we shall retain Pierce's notation for the present purpose, and we use, therefore, the definition

$$\left(\frac{E_z^2}{\beta^2 P}\right)^{1/3} = (2K)^{1/3} = (\beta/\beta_0)^{1/3}(\gamma/\beta)^{4/3}F(\gamma a).$$

Pierce has shown that, in the range of $0{\cdot}5 < \gamma a < 8{\cdot}0$, $F(\gamma a)$ is sufficiently accurately given by

$$F(\gamma a) = 7{\cdot}154 \exp(-\tfrac{3}{8}\gamma a). \qquad \ldots\ldots (7.15)$$

It will be noted that this expression gives the axial field strength on the axis; off the axis the field is generally stronger, and for a thick beam (7.15) could be a poor approximation.

The theory of propagation along a helix has been confirmed very satisfactorily by Cutler (1948), who carried out measurements on a scale model at a frequency of 560 Mc/s. He established that a uniform helix does not radiate in a radial direction, though when the helix is abruptly terminated energy is radiated strongly in the axial direction, and little is reflected. The field distribution was found to accord very well with that predicted theoretically, while the actual strength was somewhat lower. The phenomenon of dispersion, i.e. the variation of phase velocity with frequency when $\beta_0 a \cot\psi < 1{\cdot}5$, was found to be accurately described by the analysis.

It is difficult to give a simple physical explanation of why dispersion exists in helices; perhaps the question is put wrongly, and one should rather ask: why is there substantially no dispersion in the helix when $\beta_0 a \cot\psi > 3$? In ordinary waveguides of the hollow-tube kind dispersion becomes strong as the frequency is lowered and approaches the cut-off frequency. There is no cut-off frequency in a helix.

Although opinions vary considerably among workers in the field as to the elegance, or rigour, of the various theories of the travelling-wave tube, there is no doubt that Pierce's paper, on which most of the material in this section is based, is by far the most universally useful and comprehensive contribution. Considering the enormous step forward of the practical development of the travelling-wave tube associated with Pierce's name, this is perhaps not a matter for surprise. Much further work has been done by Pierce since the publication of this original paper, mainly in connection with the theory of the travelling-wave tube, but covering also a wider field, including the theory of electron-wave tubes, etc. Most of this work, and other hitherto unpublished work, have been collected and published in the book *Traveling-Wave Tubes* (Pierce 1950). It contains more of the results of original research than the published papers of all the other workers in this field together, and its usefulness to the research worker is enormous. It should be consulted by those wishing to go deeper into the subject.

§8. FURTHER DEVELOPMENTS OF THE HELIX-TYPE TRAVELLING-WAVE TUBE

The foundations of travelling-wave tube theory and practice have been laid and further development followed in the form of relatively minor modifications. When they are taken together they do, however, amount to a major advance.

8.1. *Localized Attenuation*

When the attenuation in the helix is localized somewhere along the middle section of the tube, substantially higher gain can be obtained from a travelling-wave tube than when the same attenuation is distributed uniformly. Indeed, it can be shown that such loss cannot be high enough, provided it is ensured that signal power is not reflected thereby.

The travelling-wave tube is thus effectively split into two sections. One is the input, or buncher section, the other performs a function in some ways analogous to that of the catcher in the klystron amplifier, and feedback from output to input is eliminated. In practice the localized attenuation is sometimes obtained by coating the rods (ceramic or glass) which support the helix with a layer of graphite having relatively low resistance along a length of a few centimetres. It is helpful to grade the resistance of the layer in order to minimize reflections; considering the localized lossy section of the helix as a perfect sink of power, it is apparent that the tube cannot oscillate under any circumstances, however bad the mismatches at the input and output ends may be.

The analysis of the travelling-wave tube with localized attenuation can be found in Pierce (1950, chap. 9) and will not be reproduced here. There it is shown that, provided the localized loss is put at least $N_1=(4C)^{-1}$ wavelengths beyond the input, the overall gain is only reduced by 3·5 db as compared with a hypothetical tube having no loss. (Such a lossless tube, of course, would be highly unstable.) The introduction of localized attenuation makes it possible to use helices of highly conducting wire such as copper in the output section. This is important for high-power output and efficiency, since all the r.f. output power must flow in the last few turns of the helix.

8.2. *Optimum Proportions of Helix*

Experiment showed that many tubes oscillated, though the total attenuation in the helix exceeded the forward gain considerably. At first there was a tendency to ascribe this to the presence of a rather mysterious fast wave excited on a helix only in the presence of the beam and little affected by attenuation close to the helix. Another hypothesis put forward rested on an analysis predicting a backward wave increasing with gain (Bernier 1947 a, b).

The simplest explanation, however, is that the forward gain *did* exceed the backward loss, though not at the frequency at which gain and loss measurements were made. In fact oscillations often occurred at a frequency much lower than those for which the tube and matching circuits were designed, so that, in addition, there were usually strong reflections at input and output. In view of the fact that travelling-wave tubes are normally many wavelengths long, the phase of the feedback is usually correct for oscillations near one of the frequencies of oscillation.

Another experimental fact of relevance here is that it was often found possible to increase the gain of the travelling-wave tube by defocusing the beam, as, for instance, by reducing the value of the magnetic field below that needed to get most of the beam to the collector.

Under these circumstances it is highly probable that the helix diameter is simply too large; but in order to put this into quantitative form it is necessary to consider the gain of a travelling-wave tube as the function of frequency. In the expression for gain (7.12) we have the factor CN, which contains the frequency. When the phase velocity in the beam is much slower than $\mathbf{c}$ we can put $\gamma \simeq \beta = \omega/u_0$, and we can write $CN \simeq (I_0K/4V_0)^{1/3}N$. But $K^{1/3} \simeq (\beta/2\beta_0)^{1/3}F(\gamma a)$ and $N = \beta l/2\pi$, where l is the active length of the tube. We can therefore write

$$CN = (l/2\pi a)(I_0\beta/8V_0\beta_0)^{1/3}\beta a F(\gamma a) \propto \gamma a F(\gamma a).$$

Recalling that $F(\gamma a) = 7{\cdot}154\exp(-\frac{2}{3}\gamma a)$, we obtain, therefore, the maximum gain when $\partial(CN)/\partial(\gamma a) = \exp(-\frac{2}{3}\gamma a) - \frac{2}{3}\gamma a\exp(-\frac{2}{3}\gamma a) = 0$, or $(\gamma a)_{\text{opt}} = 3/2$.

As far as the writer is aware this was first pointed out and acted upon by Rogers (1949) with apparent success. He describes the design, construction and performance of travelling-wave tubes which were developed for repeater-amplifiers in the 6 to 8 centimetre waveband.* A typical tube would give a gain of more than 20 db, varying by less than $\pm$ 1·5 db over a band of 1 400 Mc/s. The mean radius of the helix was 0·131 cm, the beam voltage 1 400 volts, and the beam current 1·5 mA, and γa is very close to 1·5 in the centre of the frequency range.

Pierce (1950, chap. 3) gives curves of the parameter $\gamma a F(\gamma a)$ for beams of various diameters from which one may deduce, for instance, that with a beam-to-helix diameter ratio of 0·9, γa_{opt} lies somewhere around 4·0.

To assume, as Rogers does, that the beam is axial and of infinitesimal cross section, is, however, not quite as unrealistic as it may sound. In any practical tube the current density in the beam will be highest near the axis; moreover, to prevent current interception by the helix, the beam diameter is made considerably smaller than the inside diameter of the helix. Further, one has to take the finite thickness of the helix wire into account, and it appears that an effective beam-to-helix ratio of 0·5 or less is quite justified. It will be seen from Pierce's curves that γa_{opt} then lies very close to 1·5.

8.3. *Dispersive Travelling-Wave Tubes*

In some applications of travelling-wave tubes a very wide bandwidth is not required, and it becomes possible to make use of the feature of *dispersion* to restrict deliberately the bandwidth over which gain is obtained (Robinson 1951). Since travelling-wave tube gain is only obtained over a restricted range of ratio of electron velocity to phase velocity, a value of γa of about 1·2 reduces the electronic bandwidth to about 10% of the centre frequency. Over such a narrow band it is then feasible to match the helix to input and output rather well. Assuming that the reflection coefficient at either end is r, and that the backward loss is L decibels and forward gain G decibels, the tube will be stable up to a maximum gain $G_{\max} = L - 20\log_{10} r$. Robinson describes tubes in which no loss has been introduced deliberately, the main part of the loss of about 13 db being due to the glass tube in which the helix is supported. With a beam current of about 100 μA stable gain of about 20 db was obtained over a band of 300 Mc/s.

Apart from the possibility of such a tube as an amplifier 'tunable' merely by changing the beam voltage, its main advantage is to be found in the extremely low beam current, which enables very good beam transmission to be obtained. This is important for low-noise travelling-wave tubes.

* These tubes were developed at the laboratories of Standard Telephones and Cables Ltd., Ilminster, Somerset.

8.4. *Low-noise Travelling-Wave Tubes*

It had been observed very early on that partition noise makes a great deal of difference to the noise factor of a travelling-wave tube (Kompfner 1947 b); a strong correlation was found to exist between the fraction of beam current intercepted and the noise figure, and the best noise figure (11 db) was obtained when only a few per cent of the total cathode current were lost on the way to the collector.

In the later stages of the development of the travelling-wave tube the main emphasis was on gain, bandwidth and output power, and an interception of 20% of the total cathode current was considered of no importance, provided the beam current was high enough to give sufficient gain to overcome the attenuation put deliberately into the path of the signal. It was therefore not surprising that noise factors between 25 and 30 db were generally reported.

The invention of the 'dispersive' travelling-wave tube (§8.3) led to a re-examination of the travelling-wave tube as a low-noise r.f. amplifier, and work was undertaken, initially, at the Services Electronics Research Laboratory, Baldock, England, to determine the limits which might be reached in practice (Robinson and Kompfner 1951).

It was possible to establish that attenuation in the helix has only a negligible influence on the noise factor. The main factor is partition noise, and most of the effort went towards reducing this. There are several distinguishable sources of partition noise:

(*a*) *Direct interception.* This phenomenon is well known at lower frequencies, and it can be shown that it may be of equal importance at microwave frequencies. In the limit it appears that if a small fraction α of the beam current is intercepted, the effective value of Γ^2 (the so-called space-charge smoothing factor) is simply equal to α. Thus to make Γ^2 much less than 0·01 less than 1% of the beam must be intercepted.

(*b*) *Non-uniform coupling of beam to the circuit* ('*induced partition noise*'). The axial field in a helix varies as $I_0(\gamma r)$ across the helix, and the strength of coupling of an individual electron to the circuit is therefore dependent on its distance from the axis. On any view of the mechanism of space-charge smoothing it follows that this phenomenon will give rise to partition noise, and it can be shown that the effective space-charge smoothing factor due to this cause is given by

$$\Gamma_i^2 = \left(\frac{\gamma b}{2}\right)^2 \left\{ \left(\frac{I_0(\gamma b)}{I_1(\gamma b)} \right)^2 - 1 \right\} - 1.$$

b is the beam radius and γ the radial propagation constant. A value of γb of 1·5 gives rise to a value of Γ_i^2 of about 0·05.

(*c*) *Velocity spread arising from non-parallel trajectories due to electron-optical defects.* In any real travelling-wave tube electron trajectories are bound to be non-parallel for a number of reasons such as misalignment of the tube in the magnetic field, non-homogeneity of the magnetic field and inevitable electron-optical defects in the electron gun.

Suppose that an electron makes an angle ϵ with the axis, proportional to its distance from the axis, $\epsilon = \epsilon_0 r/b$, and that all electrons have the same initial energy. (This neglects any effects of space charge.) It can then be shown that the space-charge smoothing factor due to this effect is given by $\Gamma_e^2 = \epsilon_0^4/48C^2$. In practice C lies around 0·02, and if Γ_e^2 is to be less than 0·005, ϵ_0 must be less than 0·1 radian, i.e. less than 5°.

(*d*) *Maxwellian distribution of initial velocities.* The Maxwellian distribution of initial velocities has been studied by MacDonald (1949 b) and was found to be unimportant in practice. Another source of noise considered was the thermal ('Johnson') noise originating in the resistance of the helix, or in whatever causes the distributed attenuation along the helix. It was shown that this will only become important when tubes with noise factors well below 10 db are considered.

Carefully avoiding all causes of partition noise, the experimental work led to the construction of tubes with noise factors around 11 db, with gain and bandwidth similar to that obtained from the normal kind of tube. A typical tube had a helix of 2 mm internal diameter, 30 cm long, with a distributed loss of 18 db. The beam voltage was 850 volts, and with a current of 165 μA a net gain of 20 db with a noise factor of 11 db was obtained. The cathode diameter was 0·5 mm and the amount of current intercepted 2 μA.

Similar results have been obtained at Stanford University (Field 1950).

Thus it has been shown that, in principle, the travelling-wave tube could be made into an amplifier with a sensitivity at least equal to, if not better than, any other practical microwave tube. It was realized, however, that partition noise is only a secondary phenomenon; the primary phenomenon is space-charge smoothing, and at that time this was a subject little understood at microwave frequencies.

8.5. *Space-charge Smoothing at Microwave Frequencies*

For a long time the only attempt to describe the process of space-charge smoothing of shot noise at very high frequencies was that by Rack (1938). On the assumption that the spread of transit times of electrons due to thermal velocities of emission is small compared with a cycle, he shows that the electron flow at the potential minimum (which has a Maxwellian distribution of velocities) can be replaced by an electron flow having a single-valued velocity, fluctuating with a mean square value equal to that of the Maxwellian distribution. Once this is accepted, one can then use the Benham–Llewellyn–Müller theory of electron dynamics in diodes, and it is found that considerable smoothing of the shot noise fluctuations can be expected. MacDonald (1950) has shown, however, that the basic assumption made by Rack does *not* apply to electron guns as used for travelling-wave tubes and klystrons at microwave frequencies; the essential part of such guns is the cathode–anode space, for which the space-charge limited diode is a good model.

The beam emerging from the electron gun can be considered as having been velocity-modulated with the equivalent velocity fluctuation as given by Rack, and modified by the space charge in the diode part of the gun. Now it has been shown (Ramo 1939 a, b) that the phenomenon of velocity modulation and bunching in an electron beam when space-charge repulsion is taken into account can be described by a system of space-charge waves. Normally there are only two waves present, one slightly faster and one slightly slower than the mean electron velocity. The interference between these two waves resembles that between longitudinal standing waves in an elastic rod which is moving with the mean electron velocity. Viewed by a stationary observer there are, along the beam regions in which there is a strong velocity modulation and no charge-density modulation, alternating with regions in which there is no velocity modulation but strong charge-density modulation. This standing wave of velocity, or charge-density modulation, is characterized by a wavelength dependent on the beam

voltage, beam current density and the geometry of the beam and surroundings and is called the 'plasma wavelength'.

In view of the doubt cast on Rack's assumption by MacDonald's work it was therefore rather surprising to find that there is good qualitative and quantitative agreement between a theory of noise in travelling-wave tubes based on Rack and Benham–Llewellyn–Müller (Pierce 1950, chap. 10) and an experiment carried out by Cutler and Quate (1950). In this experiment the shot noise in a long electron beam, produced in a conventional type of electron gun, was sampled by means of a 4000 Mc/s resonator which could be moved along the beam. Not only was the existence of the plasma standing wave of noise current confirmed, its length being exactly what it should be, but the shot noise amplitude agreed closely with the calculated value, 11 db below $2eI_0$. The noise current minima were about 20 db below shot noise; in theory there should be zero current at these points if it were not for the inevitable partition noise.

It looks, therefore, as if Pierce's theory of noise in travelling-wave tubes, though built on an untenable assumption, gives the correct result. The solution of this dilemma has been given by Robinson (1952) on the basis of a careful analysis of electron flow in the diode part of the electron gun, and it appears that Pierce's theory, rather fortuitously, gave the right answer. Had a much higher frequency or a longer cathode–anode distance been used there would have been considerable discrepancies.

In view of this rather unsettled position it is unprofitable to proceed further. It is obvious, however, that any progress made in the understanding of shot noise in electron beams at microwave frequencies will have an important bearing on the noise factors obtainable from travelling-wave tubes.

8.6. *Electron Optics in the Travelling-Wave Tube*

Electron beams having currents of tens of milliamperes at a few thousand volts and diameters of a few millimetres were at first produced in electron guns of the so-called 'Pierce'-type (Pierce 1940, Field 1946) either with planar or concave spherical cathodes. A uniform magnetic field produced by a solenoid kept the beam together once it had been formed, and no special difficulty was experienced in projecting such beams over lengths of up to 50 centimetres. The gun itself was often kept in the uniform part of the solenoidal part, though it had been suggested the gun be kept in the fringing part of the solenoidal field in order to make the electrons follow the convergent lines of force.

It was soon evident that merely increasing the beam current did not always lead to an increase in the gain or r.f. power output from a travelling-wave tube. Indeed, often an increase of the current beyond a certain limiting value led to a decrease of gain.

On the assumption that all electrons move in straight parallel trajectories it can be shown to a first approximation that there will be a depression of potential due to space charge: $\Delta V = -1{\cdot}5 \times 10^4 I_0 V_0^{-1/2}(r/b)^2$, where b is the beam radius.

Thus the electrons on the outside of the beam will move faster than the inside ones, and the travelling-wave interaction between beam and wave will exhibit interference symptoms and become ineffective. This then seemed a serious limitation in the development of the travelling-wave tube for some time.

Important progress was made when it was realized that a type of electron flow can be obtained in which space-charge repulsion and centrifugal force

are exactly balanced by the force due to the magnetic field (Brillouin 1945) and that this would lead to electron flow with uniform axial velocities (Samuel 1949, Pierce 1949, Wang 1950).

A sketch of an arrangement designed to give 'Brillouin'-type electron flow is given in fig. 13.

The best conditions are obtained when the gun is in a field-free space and when the initially convergent electron stream enters the magnetic field abruptly with parallel trajectories. This can be done, as shown, by shrouding the solenoid in a thick iron shield with an aperture on the axis where the beam is injected. At this point spiralling of the electrons due to the radial component of magnetic field is set up such that when the current density, beam potential and magnetic field are correctly related to each other, the beam as a whole rotates with an angular velocity $\dot{\theta} = -eB/2m$. The axial velocity is then the same for all electrons: $u_0 = (2eV_0/m)^{1/2}$, where V_0 is the potential on the axis.

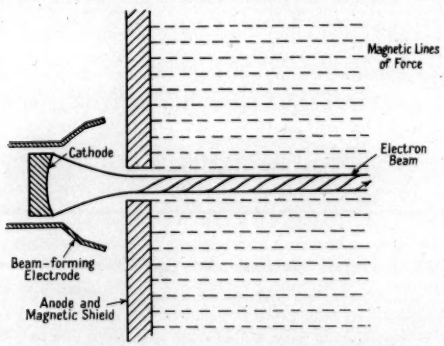


Fig. 13. Brillouin-flow type of electron gun ('Wang-gun').

The total beam current I_0, the beam radius b, the magnetic field B, and V_0 are then related by

$$I_0 = \frac{\pi\epsilon}{\sqrt{2}}\left(\frac{e}{m}\right)^{3/2} B^2 V_0^{1/2} b^2 = 1{\cdot}45 \times 10^6 B^2 V_0^{1/2} b^2$$

when the beam is completely filling a conducting tube at a potential V_b with respect to the cathode. ΔV, the potential drop between the outermost electrons and the axis, is given by

$$\Delta V = V_b - V_0 = \frac{e}{8m} B^2 b^2 = 1{\cdot}107 \times 10^{-12} B^2 b^2.$$

Substituting $V_0 = V_b - \Delta V$, and keeping V_b constant, one sees that there will be a value of B^2b^2 for which the beam current is a maximum. This happens when $\Delta V = \frac{2}{3}V_b$ and gives

$$I_{\max} = \frac{16\pi\epsilon}{3\sqrt{6}}\left(\frac{e}{m}\right)^{1/2} V_b^{3/2} = 25{\cdot}4 \times 10^{-6} V_b^{3/2};$$

that is to say, I_{max} depends only on the applied beam voltage. For every given value of b there is, however, only one correct value of B, and to produce a very thin beam may require a very high magnetic field.

Under these limiting conditions the potential drop between innermost and outermost electrons is exactly twice the potential of the innermost electrons, and the axial velocity of all electrons is therefore $3^{-1/2}=0{\cdot}578$ of the velocity of electrons that would move rectilinearly in a potential V_b.

One half of the total beam energy is then rotational and is, therefore, not available for conversion into r.f. energy if the r.f. fields are mainly axial (as they usually are in klystrons and travelling-wave tubes).

Normally one will not use Brillouin flow under these extreme limiting conditions, as 'mild' Brillouin flow already gives quite striking results.

Field (1950) describes a helix-type travelling-wave tube working in the 2 000 to 4 000 Mc/s region with which 60 watt r.f. output power was obtained. Using Brillouin flow he was able to project a beam of 200 mA at 3 000 volts down a helix with a bore of 2·8 mm and 33 cm long, with a loss of only one milliampere. The feature which is important for good efficiency and high gain is, of course, the fact that all electrons have practically the same axial velocity, regardless of their potential.

Another type of tube, already commercially available, and presumably utilizing Brillouin-flow focusing, has recently been described (Bryant 1950). It is designed to work in the 4 400 to 5 000 Mc/s band, to give 10 watt r.f. output power with 20 db gain, and has a maximum efficiency of 20%, agreeing closely with a theoretical prediction. The limit to the power output is set by heating of the copper-plated helix near the output by the r.f. currents flowing in it.

8.7. *The Helix-type Travelling-wave Tube carried to its Limits*

By using gradually tapered sections of helix the bandwidth of the circuit match has been improved considerably (Lund 1950, Field 1950). Field describes a travelling-wave tube giving from 40 to 50 db gain from 300 to 1 000 Mc/s; at 500 Mc/s 26% efficiency was measured, and with reduced collector potential 50%.

A power amplifier has been described (Webber 1950) which, at a frequency of 450 Mc/s, gives a gain of 25 db at 20% efficiency. The output power is 1 000 watts and the helix is formed by a water-cooled copper tube. The limit here is due to r.f. heating of the glass envelope. It can be expected that considerable progress will be made with tubes of this kind since in the United States there is a television frequency band in this region.

At the other end of the scale is an attempt, made at Bell Telephone Laboratories in the United States (Little 1951), to make a travelling-wave tube amplify at a wavelength of 6 mm. Using a 0·075 mm tungsten wire wound on a 0·75 mm mandril with 0·165 mm pitch, a gain of 3 db was obtained at a beam voltage of 1 000 volts and 1 mA beam current. A magnetic field of several thousand gauss was used to ensure rectilinear electron paths. It cannot be said that the outlook for the helix-type travelling-wave tube at these wavelengths is good, and it will be shown later that there are other types of travelling-wave tube that appear more promising. However, a type of tube that works from 300 to 50 000 Mc/s covers an enormous frequency range, as wide as any other type of tube yet envisaged.

§ 9. STRUCTURES OTHER THAN HELICAL WHICH SUPPORT SLOW WAVES

9.1. *Fundamental Mode of Operation*

It is apparent that a wire helix cannot dissipate much heat. Compared, for instance, with the solid copper block out of which magnetron anodes are made, or with the water-cooled copper cylinder of high-power transmitter triodes, the helix of a travelling-wave tube is thin and fragile, and many travelling-wave tubes have been destroyed when rather more than usual of the electron beam, by some mischance, hit the helix. Therefore a search has been going on since the early days for structures capable of supporting a slow wave and producing a field configuration suitable for interaction with an electron beam. Some of these are: the corrugated waveguide used in many linear accelerators (Slater 1946, Chu and Hansen 1947, Fry and Walkinshaw 1949, Brillouin 1948), a waveguide loaded with dielectric material of high dielectric constant (Frankel 1947, Banos, Saxon and Gruen 1951), a waveguide coiled into a helix, with the inside wall removed; in fact, the number of possible configurations is very large and we shall not attempt to enumerate them all. We shall only discuss some general principles and some structures that have been of importance in practice.

If a structure is to be suitable for use with a travelling-wave tube we require it to have the following properties: (i) it should have a strong axial field for a given power input, that is to say, it should have a high effective impedance; (ii) its phase velocity must not vary much with frequency, in order not to restrict the bandwidth over which amplification can be obtained at one particular beam velocity; (iii) its attenuation should not be too high, in order not to lead to a serious loss of gain and overall efficiency; (iv) it should enable interaction with a beam of the largest possible cross section to be obtained; (v) it is an advantage if heat is easily conducted away from the areas that might be bombarded by the electron beam; (vi) the elements of which the structure is built should be as large as possible and should not require excessive dimensional accuracy; (vii) it should not be difficult to build.

Some of these requirements are obviously not amenable to quantitative treatment, and comparisons will depend a great deal on common sense and experience. Others can be dealt with analytically, and it will be seen that some requirements are conflicting.

K, the effective impedance, can be written $K=E_z^2/2\beta^2 P$, where E_z is the axial field strength, $\beta=\omega/v$ is the propagation constant, P is the power flow. We can express the power flow in terms of stored energy per unit length W and group velocity v_g: $P=Wv_g$, and thus $K=E_z^2/2\beta^2 Wv_g$.

The group velocity is given by

$$v_g=v\left(1-\frac{\omega}{v}\frac{\partial v}{\partial\omega}\right)^{-1},$$

and we note that small dispersion means small $\partial v/\partial\omega$. Thus it is necessary to have $v_g\simeq v$ for a broad band. On the other hand, near a cut-off $v_g\to 0$, $K\to\infty$ and $\partial v/\partial\omega\to\infty$, that is to say we approach infinite impedance with zero bandwidth.

If the structure has a quality factor Q, it can be shown that the attenuation constant α (amplitude along structure as $\exp(-\alpha z)$) is given by $\alpha=\omega/2Qv_g$. Thus decreasing v_g, which gives a higher impedance, increases the attenuation.

Field (1949) has compared the helical waveguide structure mentioned earlier with a baffle-loaded cylindrical waveguide of similar dimensions. The field configurations are very similar and both have the same stored energy for a given field strength. He found that the product of impedance and bandwidth was approximately the same in both cases.

Pierce (1950, chap. 5) has compared a helix, a structure composed of shallow pill-box resonators, and an array of thin wires half a wavelength long between two infinite conducting planes. The impedance of a helix has been given in §7.5. The corresponding impedance of the resonators (their coupling adjusted to give a band as broad as that of the helix) is given by

$$(E^2/\beta^2 P)^{1/3} = 5{\cdot}36(v/\mathbf{c})^{1/3}(v/v_g)^{1/3}$$

and the corresponding impedance of the array of half-wave wires

$$(E^2/\beta^2 P)^{1/3} = 6{\cdot}20(v/v_g)^{1/3}.$$

It should be noted that this structure cannot propagate a wave unless additional coupling (such as might be provided by an electron beam) is present. It appears that the helix is a very good circuit from the points of view of requirements (i) (ii), (iii) and (vii).

One structure, on which much theoretical and experimental work has been done, is the baffle-loaded coaxial line or, as sometimes called, the 'disc-on-rod'

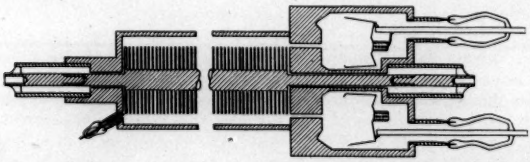


Fig. 14. Cross-section of baffle-loaded coaxial line tube built at Stanford University.

structure, as shown in fig. 14 (Field 1950). The axial field which is available for interaction with the electrons is now distributed around the circumference of the discs and enables a hollow, or annular, electron beam to be used which may have a cross-sectional area many times larger than used in the helix-type travelling-wave tube. Thus we can expect more current and more power. The phase velocity will depend on frequency, hence this type of tube has a narrow bandwidth over which it amplifies, but correspondingly the impedance is high.

Calculations of the phase velocity with and without the beam and of the gain to be expected from such a tube have been compared with experimental tubes (Dewey, Parzen and Marchese 1951), and it appears that agreement is satisfactory. A gain of about 20 db was obtained at 4 600 Mc/s with a beam voltage of 4 000 volts and beam current of 200 mA.* The best values of efficiency and power that have been obtained so far are 7% and 125 watts at about 4 500 volts and 380 mA. The bandwidth is of the order of 100 Mc/s. It seems that higher power and efficiency should be obtainable from tubes of this general kind in

* The development of this tube has been carried out at Federal Telecommunication Laboratories, Nutley, N.J., U.S.A.

future. It may be remarked, however, that such tubes are certainly not easy to construct.

One of the chief constructional difficulties is the accuracy required in the dimensions. There is usually, in the repetitive kind of structure discussed above, at least one dimension which must not vary by more than, say, a few parts in a thousand. This applies to all resonant slots—and to get high impedances slots have to be resonant, i.e. one-quarter of a wavelength long, for instance. Thus, if the structure is to be fabricated from parts and soldered or brazed together, the amount of solder projecting beyond the joints must be carefully controlled. If the structure is machined out of solid metal this limits the shortest wavelength at which it may be used. A fault or inaccuracy in only one of, say, a few hundred slots will cause reflection of power that may seriously affect the performance. On the other hand the construction techniques required are very similar to those used in the manufacture of magnetrons, which are made in very great numbers.

9.2. *Harmonic Mode of Operation*

Recently attention has been directed to the possibility of obtaining interaction between the electron stream and a wave which travels considerably faster than the electrons, by making use of the space-harmonics of the wave.

Suppose we have a structure in which there are slots separated by a distance d in the z-direction and a wave travelling along this structure with a phase velocity v. The phase displacement between adjacent slots is then $\theta = \omega d/v$, where $\omega/v = 2\pi/\lambda$; thus an electron would interact with the wave if it travelled with a velocity $u_0 = v = \omega d/\theta$.

We assume in the first place that electrons are exposed to the r.f. field only while they are crossing a slot. Suppose an electron travelled much more slowly than the wave but crossed every slot at an instant when the phase of the wave at the slot is the same; then the time taken by an electron to cover the distance d must equal the time taken by the wave to travel d plus an integral number of wavelengths, i.e. $d/u_0 = (n\lambda + d)/v$, whence $u_0 = v\{(2\pi n/\theta) + 1\}^{-1}$. For $n = 0$, $u_0 = v$, giving the fundamental mode of operation; $n = 1$ represents the first harmonic (or 'Hartree') mode. Higher values of n are not normally of interest.

For a wave travelling in the direction opposite to that of the beam $u_0 = v\{(2\pi n/\theta) - 1\}^{-1}$. In both cases we observe that the electron speed for synchronism with the harmonic of the wave can be considerably less than the phase velocity of the wave; to achieve this object however it is now not necessary to have near-resonant elements or slots. All that is required is that an electron should be shielded from the wave periodically, as when passing through a field-free hole or longitudinal slot, and exposed to an r.f. field periodically, as when crossing a transverse slot across which an r.f. field exists. The most important result of this technique is the considerable reduction in beam velocity for a given physical separation of the circuit elements. This has made it possible to build tubes that work in the 6 mm wavelength region (Millman 1951).

Figure 15 gives diagrammatic longitudinal and cross sections of the so-called 'spatial' harmonic travelling-wave amplifier developed at Bell Telephone Laboratories. The structure consists essentially of a waveguide loaded by a longitudinal ridge which has a number of transverse slots. Without the slots the phase velocity in the ridge waveguide would be close to c over a wide range of frequencies. With the slots the phase velocity will be reduced, though still too

far high for direct interaction with a low-voltage electron beam, i.e. in the fundamental mode. Operating the tube in the first harmonic mode, the electron beam voltage was 1 220 at a wavelength of 6·25 mm; the 100 slots were spaced 0·534 mm between centres, 0·178 mm wide and seven-eighths of a wavelength deep. The fundamental phase velocity is adjusted to give relatively broad-band operation by making $\partial u_0/\partial\omega = 0$ at the operating point.

In the early experiments on this type of tube the beam was sent close to the slots and it was hoped that the axial component of the fringing field of the slots would be sufficient. This, however, was not so, and therefore three longitudinal slots 0·25 mm wide were cut into the structure. When the whole structure is flooded with an electron beam, these slots enable a far more intimate coupling

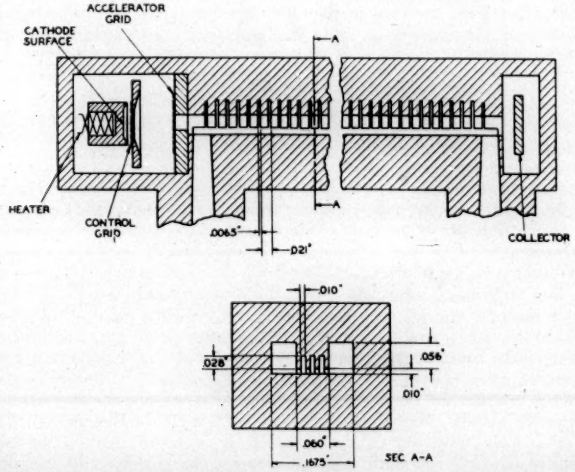


Fig. 15. Schematic diagram of space-harmonic amplifier according to S. Millman, which gives about 18 db gain in the 6-mm wavelength region.

between beam and fields to be obtained. Care is taken to make the electron trajectories as parallel as possible, and a magnetic field of about 1 600 gauss is required to hold the beam together.

Tapered waveguides are used to feed into and out of the structure with the help of matching sections, and a fairly good match is obtained over a band of about 10% of the centre frequency. The circuit attenuation is about 11 db.

Typical performance of the tube is:

beam voltage: 1 200 volts; beam current: 4 mA; maximum r.f. power output: 20 mW; power gain: 20 db; bandwidth: 3% of 50 000 Mc/s, = 1 500 Mc/s.

It will be realized that this performance is quite an achievement and opens up the possibility of exploring still shorter wavelengths.

Somewhat different types of space harmonic tubes are being developed at Stanford University, California, U.S.A.* Diagrammatic sketches are shown in fig. 16 (*a*) and (*b*).

In the tube shown in (*a*) both axial and transverse harmonic components of the r.f. field are acting on the beam, in the tube shown in (*b*) the component is mainly axial. The main difference between this type of tube and Millman's is that here the electrons are at no time screened from the fundamental components of the field. Although the time average of the latter is zero, they are bound to have some disturbing effect. The structure is fundamentally a rectangular waveguide propagating a transverse electric wave (H_{01}-wave), loaded in the one case with transverse ridges, in the other with vertical posts. One advantage of this construction is that the tubes may be designed to operate near the low-frequency cut-off (in which the critical dimension is the width of the guide) so that a low

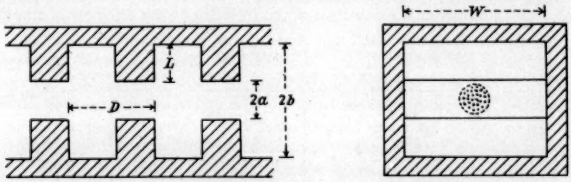


Fig. 16 (*a*). Space-harmonic travelling-wave tube in which both axial and transverse harmonic components of the radio-frequency field are acting on the electron beam.

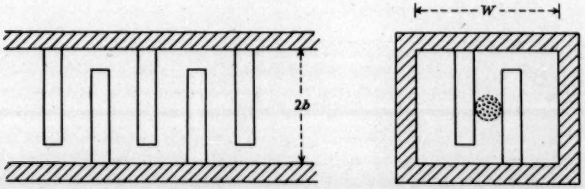


Fig. 16 (*b*). Space-harmonic travelling-wave tube in which mainly the axial harmonic component of the radio-frequency field acts on the electron beam.

group velocity and high impedance are obtained. (In the fundamental mode of operation high impedance is commonly obtained by operating near the high-frequency cut-off, in which case the critical dimension is the depth of the slots.)

The tubes were designed to operate near 3 cm wavelength, and oscillations have been observed in this range with beam voltages around 2 000 and currents of several milliamperes. No data are yet available of the performances of these tubes as amplifiers, though there is little doubt that they will be able to amplify as efficiently as the more common types of travelling-wave tube; furthermore, these structures are bigger and more robust than any operating in the fundamental mode and are therefore eminently fitted to be scaled down to much shorter wavelengths.

* The writer is indebted to Professor L. M. Field for information about these tubes.

§ 10. TRAVELLING-WAVE INTERACTION IN CROSSED ELECTRIC AND MAGNETIC FIELDS

The most serious limitation of the travelling-wave tube is its relatively low efficiency. It can be argued that a device which depends for its function on the equality of the velocities of beam and wave cannot be efficient since, as soon as an appreciable amount of energy has been transferred from the beam to the wave, the beam will be too slow to interact with the wave. In actual fact the difference between the velocities of beam and wave which can be permitted without impairing the interaction process may be quite large, and hence efficiencies between 20 and 30% are possible and have been attained. On the other hand, the theoretical efficiency of *magnetrons* approaches 100% in the limit, and 70 to 80% has been attained in practice. It is inviting therefore to examine the possibility of applying the magnetron principle to the travelling-wave tube. This was very likely the kind of consideration which led Lerbs (Warnecke, Kleen, Lerbs, Doehler and Huber 1950) to propose the principle of the travelling-wave magnetron amplifier.

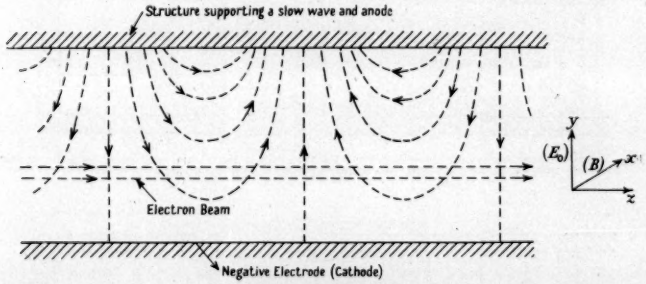


Fig. 17. Diagram of electric lines of force of the radio-frequency field in the travelling-wave magnetron amplifier.

The reason for the high efficiency of the magnetron is that electrons are made to give up *potential* energy to the r.f. field while retaining their *kinetic* energy. Thus their velocity stays practically constant and *equal* to the phase velocity of the wave while they move from a region of high potential to a region of low potential, and the electronic efficiency (for small signals) is given by the initial potential energy divided by the total d.c. energy of an electron. In order to realize this kind of trajectory we have set up crossed electric and magnetic fields. This is done for instance by applying a d.c. potential difference in the space between two parallel planes, say the cathode plane and the anode plane, producing a uniform field E_0 in the y direction, while a uniform magnetic field is applied in the x direction, parallel to the planes. Provided the magnetic field is strong enough ('beyond cut-off') electrons emitted from the cathode, for instance, will describe trochoidal paths, with a mean velocity in the z direction given by $u_0 = E_0/B$. Alternatively, and preferably, we can inject electrons with a velocity u_0 into the space parallel to the z direction, and such electrons will, in the absence of r.f. field, continue to move rectilinearly in the z direction.

Consider the configuration shown in fig. 17. There a thin electron beam is injected in the z direction with the velocity $u_0 = E_0/B$ near to the lower planar

electrode. We have also a wave travelling with the same velocity u_0 in the same direction, as indicated by the electric lines of force. The r.f. field will disturb the trajectories as explained below.

Take an electron that enters the interaction space with the correct velocity u_0 for rectilinear propagation at an instant when the r.f. field E is purely transverse and in the same direction as the d.c. field. This electron will then be too slow for rectilinear propagation in a system of crossed fields $(E_0+E)/B$ and will describe a trochoidal path, but with a drift velocity slightly higher than u_0. It will therefore drift into a region where there is a retarding component of r.f. field. To see what effect a retarding field has on an electron, let us consider an electron entering the interaction space at an instant when the r.f. field is purely retarding. Such an electron will transfer energy to the wave, that is to say lose potential energy, and must move closer to the anode. On the other hand, an electron in a purely accelerating field will absorb energy from the wave and move closer to the cathode. An electron entering at a time when the field is purely transverse and in opposition to the d.c. field will be forced into a trochoidal path, but with a drift velocity slightly lower than u_0, and will therefore also drift into a region where the field is retarding.

It will be seen that the function of the transverse components of the r.f. field is to bunch the electrons and to drive them into the retarding portion of the field. The function of the retarding component of the field is to transfer potential energy from the electrons to the field, the electrons moving closer to the anode. The accelerating component of the field drives electrons towards the cathode, where they are lost from the interaction. If the beam had originally been injected close to the cathode little energy is lost through these electrons, while almost all the potential energy of the rest of the beam is given up to the r.f. field: thus the high efficiency is explained.

When the electron beam originates at the cathode with practically zero emission velocity matters are less favourable, since the beam has now extra energy in the form of rotational energy of the cycloidal motion. This energy is normally not available for conversion into r.f. energy. Moreover, because of the amplitude of the excursion of electrons in the y direction, the cathode–anode distance has to be increased, which in turn requires an increase in the potential difference across it in order to maintain the same value of E_0.

Regardless of whether the initial motion of the electrons is rectilinear or cycloidal, one can thus arrive at a crude estimate of the efficiency: $\eta=(V_a-V_0)/V_a$, where V_a is the anode potential and V_0 the highest potential reached by the beam in the absence of r.f. power.

Small-signal theories of the travelling-wave magnetron amplifier have been given by the French workers Brossart and Doehler (1948) and by Pierce (1950, chap. 15) and we shall not reproduce them here since they follow the pattern set by the earlier travelling-wave tube theories. It turns out that the propagation constant Γ is obtained as the solution of an equation of the fifth degree, indicating that gain may be possible in two ranges of parameters. In one range the electron velocity is near the phase velocity of the wave, in the other $\beta_e/\beta_1=1-\beta_m/\beta_1$. β_e is the electron propagation constant $=\omega/u_0$, β_1 is the circuit propagation constant $=\omega/v$, $\beta_m=\omega_m/u_0$, where $\omega_m=\boldsymbol{eB/m}$, the cyclotron radian frequency. In the latter case the wave-retarding structure would have to be at a negative potential, the electrons moving away from it during their transit.

In either case the gain of the increasing wave in decibels per wavelength is proportional to the square root of the current rather than to the cube root, as in the ordinary travelling-wave tube. This means lower gain for the same circuit and current.

Tubes operating in the first range of parameter have worked and descriptions and results have been given (Warnecke and Guenard 1948).

Both cylindrical and plane geometries have been tried and give comparable results. The wave-retarding structure is a flattened helix, matched to coaxial lines at input and output. The electron beam is produced in an electron gun and is made as near rectilinear—that is without cycloidal motion—as is possible by injecting it at a point between the positive and negative electrodes where the potential is such as to equal the kinetic energy of the beam.

Typical values for a tube with planar geometry are: wavelength 22 cm, anode voltage 2000, negative electrode -600 v, cathode 0 v, distance between electrodes 0·3 cm, active length 20 cm, electron velocity $1{\cdot}36\times10^9$ cm/sec, magnetic field 600 gauss, maximum gain 7 db, bandwidth 100 Mc/s. The tubes have to be rather precisely constructed, and output power and efficiency are very sensitive to variations in the electron optical system. With optical systems approaching the ideal, 200 watt r.f. output power and 40% efficiency have been obtained.*

Evidently considerable progress has to be made before this type of tube can compete with other types of microwave amplifiers; however, the promise of high efficiency, and also of higher power, is there, and the principle is sound.

§ 11. ELECTRON PLASMA WAVES

Before discussing the more recently discovered mechanisms of amplification of waves in electron and ion streams in the absence of any material structures alongside the electron streams, it is advisable to study the behaviour of a single electron stream under some simplifying assumptions. A qualitative description has already been given in § 8.5, and some of the quantities will now be calculated.

A fairly complete and rigorous theory of waves in electron beams, either in free space or surrounded by conductors, of cylindrical and of planar geometry has been given (Ramo 1939 a, b, MacFarlane and Woodward 1950). Here we shall use the method of Pierce as applied to an electron stream of infinite cross section moving with a mean velocity u_0 in the z direction. We shall again assume that all a.c. quantities, namely voltage V (defined by $\partial V/\partial z \equiv E_z$), charge density ρ, convection current density i and velocity v vary as $\exp(j\omega t-\Gamma z)$, where Γ is the unknown propagation constant.

It has been shown in § 7.1 that a voltage V gives rise to a velocity modulation described by

$$v=-\,e\Gamma V/mu_0(j\beta_e-\Gamma),$$

where $\beta_e=\omega/u_0$, the 'undisturbed' beam propagation constant.

It has also been shown that a voltage V gives rise to a convection current i:

$$i=j\frac{I_0\beta_e\Gamma V}{2V_0(j\beta_e-\Gamma)^2}, \qquad \ldots\ldots(11.1)$$

* Extensive studies on this type of tube have been made at the Centre de Recherches de la Compagnie générale de Télégraphie sans Fil, Paris.

where I_0 is the d.c. current density and V_0 is the d.c. beam voltage. We shall now assume that the voltage V is entirely due to the charge distribution in the beam itself.

From Poisson's equation we have, in the one-dimensional case,

$$\frac{\partial^2 V}{\partial z^2} = -\frac{\rho}{\epsilon}, \qquad \Gamma^2 V = -\frac{\rho}{\epsilon}, \qquad V = -\frac{\rho}{\epsilon\Gamma^2}.$$

From the equation of continuity

$$\frac{\partial i}{\partial z} = -\frac{\partial \rho}{\partial t}, \qquad -\Gamma i = -j\omega\rho, \qquad \rho = -j\frac{\Gamma i}{\omega}.$$

Therefore

$$V = ji/\epsilon\omega\Gamma. \qquad \ldots\ldots(11.2)$$

Equations (11.1) and (11.2) are self-consistent when $I_0\beta_e/2V_0(j\beta_e - \Gamma)^2\epsilon\omega = -1$, which yields $-\Gamma = -j\beta_e \pm j(I_0\beta_e/2V_0\epsilon\omega)^{1/2}$. This can also be written

$$-\Gamma = -j\beta_e[1 \mp \omega_p/\omega] = -j\beta_e \pm j\beta_p,$$

where ω_p is the so-called electron plasma frequency defined by

$$\omega_p^2 = \boldsymbol{e}I_0/\epsilon\boldsymbol{m}u_0, \qquad \beta_p = \omega_p/u_0.$$

Every wave is therefore composed of two separate waves, with different phase velocities, $u_1 = u_0(1 - \omega_p/\omega)^{-1}$, $u_2 = u_0(1 + \omega_p/\omega)^{-1}$, one slightly higher and one slightly lower than u_0, the mean velocity.

Suppose we write the velocity distribution

$$v = v_0 \exp(j\omega t - \Gamma z),$$

then, on the assumption that $\omega_p/\omega \ll 1$, the two velocity waves can be written $v_{1,2} = \frac{1}{2}v_0 \exp(\mp j\beta_p z) \exp(j\omega t - j\beta_e z)$ or $v = v_0 \cos(\beta_p z) \exp(j\omega t - j\beta_e z)$.

Similarly the convection current can be written

$$i = -\frac{v_0\omega}{u_0\omega_p} I_0 \sin(\beta_p z) \exp(j\omega t - j\beta_e z)$$

and the voltage $V = -2\dfrac{v_0\omega_p}{u_0\omega} V_0 \sin(\beta_p z) \exp(j\omega t - j\beta_e z)$.

The meaning of these equations is this: suppose we impress on a beam, initially of uniform current density and velocity, a velocity modulation v_0 at $z = 0$. This will be turned into a convection current i which reaches a maximum amplitude when $\sin(\beta_p z) = 1$, $\beta_p z = \pi/2$ or $z = \lambda_p/4$ (where λ_p is the so-called electron plasma wavelength). At that point the velocity modulation is zero and all the a.c. kinetic energy has gone into a.c. potential energy as indicated by the voltage V. At $z = \frac{1}{2}\lambda_p$ the energy is again all kinetic and we have pure velocity modulation; at $z = \frac{3}{4}\lambda_p$ we have another current maximum, and so forth. The description of the system when ω_p is no longer very small compared with ω is slightly more complicated, but similar in principle.

What we have described is an electron-plasma oscillation (Tonks and Langmuir 1929) taking place in a beam that is moving relative to a stationary observer. The initial disturbance is, in our case, a velocity modulation, and the charge thus disturbed will oscillate indefinitely while it is carried along with the beam. The point of interest to us, however, is the fact that consequently the electron beam behaves very much like a transmission line* or waveguide, in that it transmits

* This aspect has lately been elaborated by Mathews (1951 a, b), who based on it a complete theory of the travelling-wave tube, deriving in general the propagation properties of two coupled transmission lines in motion relative to one another.

signals (we can regard any of the a.c. quantities as a signal) without loss. The splitting of the wave is akin to excitation of a waveguide in two neighbouring modes; this sometimes happens and does not affect the principle.

It should be kept in mind that all this is only true for small signals. With large signals electrons may overtake each other, just as waves on the sea turn into breakers, and energy goes into higher harmonics of the signal frequency and is lost.

Overtaking and severe non-linearity will occur (Brillouin 1946) when $v_0/u_0 \geqslant \omega_p/\omega$. It is clear then that ω_p/ω expresses some property of an electron stream connected with its strength or stiffness, and small-signal theory only applies when v_0/u_0 is well below this limit.

§ 12. THE ELECTRON-WAVE TUBE

It was very likely the realization that an electron beam can be regarded as a special kind of transmission line capable of propagating slow electric waves that led to the idea of using such a beam instead of a helix in a travelling-wave tube, that is to say, a tube with two electron beams and no helix. Now, in the travelling-wave tube, the beam and the helix need to be in close proximity for good coupling and high gain to be obtained. Analogously, one would expect to put the beams as close together as possible; this is indeed not difficult, as one can make them coincide, that is to say interpenetrate completely, for two distinct and coinciding electron beams can exist if they are given different velocities; their charge densities may, but need not, be the same.

Some considerations such as these may have led to the invention of the two-beam tube, or electron-wave tube (Haeff 1948, 1949).*

The principle of the electron-wave tube is very simple: the signal may be impressed on one of the two beams, say in the form of a velocity modulation, by means of a resonator, helix, or any other means of coupling. This beam is then either mixed with another beam of different velocity, or made to travel in close proximity to such a beam, and the signal is amplified approximately exponentially with distance, in either beam. The amplification process can be explained in qualitative terms in a manner analogous to that employed in § 4, one beam taking the place of the transmission line. At a suitable point where the signal has been sufficiently amplified r.f. power is abstracted from one of the beams, again by means of a resonator, helix, etc. Thus the need for a long uniform circuit close to the electron beam carrying a slow wave capable of dissipating heat from stray electrons, and apt to transmit r.f. power reflected from a mismatch near the output and thus leading to self-oscillation, is eliminated. These considerations make it likely that the electron-wave tube or a tube derived from it may eventually be of particular interest at the very shortest wavelengths.

12.1. *Theory*

Let the two interpenetrating beams, of infinite cross section, be described by the following parameters: I_1, I_2 d.c. current densities, u_1, u_2 d.c. velocities, $\rho_{01} = -I_1/u_1$, $\rho_{02} = -I_2/u_2$ d.c. charge densities, V_1, V_2 d.c. beam voltages, ρ_1, ρ_2 a.c. charge densities, v_1, v_2 a.c. velocities, V a.c. potential, $\beta_1 = \omega/u_1$, $\beta_2 = \omega/u_2$. Any a.c. quantity will again be described by $\exp(j\omega t - \Gamma z)$.

* Growing waves in mixed ion and electron streams had also been discussed somewhat earlier, though no possible applications had been suggested (Pierce 1948). The idea seems to have originated also at other places (see Nergaard 1948, where a proposal by C. W. Hansell is described).

Using the method outlined in § 7.1, it can be shown that the a.c. charge density in the two beams is given respectively by

$$\rho_1 = \frac{eI_0\Gamma^2 V}{mu_0^3(j\beta_1-\Gamma)^2}, \qquad \rho_2 = \frac{eI_0\Gamma^2 V}{mu_0^3(j\beta_2-\Gamma)^2}.$$

We have made the special assumption that $I_1/u_1^3 = I_2/u_2^3 = I_0/u_0^3$. (This is equivalent to saying that the electron beams originate at electron guns obeying the Langmuir–Childs law.) I_0 may be considered a mean current density, u_0 a mean velocity $u_0 = 2u_1u_2/(u_1+u_2)$ and V_0 a mean potential defined by $u_0 = (2eV_0/m)^{1/2}$. Let us also define a mean phase constant $\beta = \omega/u_0$ and a fractional velocity separation $b = (u_1-u_2)/(u_1+u_2)$, whence $\beta_1 = \beta(1-b)$, $\beta_2 = \beta(1+b)$. Thus the total a.c. charge density ρ is given by $\rho = \rho_1 + \rho_2$,

$$\rho = \frac{I_0\Gamma^2}{2u_0V_0}\left[\frac{1}{\{j\beta(1-b)-\Gamma\}^2} + \frac{1}{\{j\beta(1+b)-\Gamma\}^2}\right]V.$$

From Poisson's equation $\partial^2V/\partial z^2 = -\rho/\epsilon$ we have, as before, $V = -\rho/\epsilon\Gamma^2$. Eliminating ρ and V, we obtain an equation for Γ:

$$\frac{1}{[j\beta(1-b)-\Gamma]^2} + \frac{1}{[j\beta(1+b)-\Gamma]^2} = \left(\frac{\omega}{\omega_p}\right)^2\frac{1}{\beta^2}, \qquad \text{.....(12.1)}$$

where $\omega_p^2 = eI_0/\epsilon mu_0$ as before. Writing $-\Gamma = -j\beta(1+h)$ in eqn. (12.1) and solving for h, we obtain

$$h = \pm b\left[\left(\frac{\omega_p}{b\omega}\right)^2 + 1 \pm \frac{\omega_p}{b\omega}\left\{\left(\frac{\omega_p}{b\omega}\right)^2 + 4\right\}^{1/2}\right]^{1/2}. \qquad \text{.....(12.2)}$$

The positive sign inside the brackets gives a real value of h, and hence four unattenuated waves. When $(\omega_p/\omega)^2 > \frac{1}{2}b^2$ the negative sign inside the brackets gives two imaginary values for h, that is, one growing and one attenuated wave. We can then maximize h with respect to b and find $b_{opt} = \sqrt{3}\omega_p/2\omega$, whence $h_{max} = \pm\frac{1}{2}j\omega_p/\omega$. Thus the growing wave can be described by

$$\exp(j\omega t - j\beta z)\exp(\tfrac{1}{2}\beta z\omega_p/\omega),$$

that is to say, a.c. quantities in the beam grow as

$$\exp(\tfrac{1}{2}\beta_p z), \qquad \text{.....(12.17)}$$

where $\beta_p = \omega_p/u_0$, the electron-plasma phase constant. This describes the maximum theoretical rate of increase of the growing wave; in any real tube the beams may not interpenetrate and will have a finite cross section which will reduce the gain.*

There is now no point in using the boundary conditions to determine the initial loss, as was done in the travelling-wave tube case. The means of coupling in and coupling out are bound to introduce an amount of loss very difficult to evaluate, so that we shall not arrive at an equation of the overall gain of the electron-wave tube. It can be shown, however, that even with very inefficient coupling the overall gain may easily exceed the best that can be obtained from the travelling-wave tube.

The bandwidth of the electronic gain can be obtained from eqn. (12.2) and turns out to be very wide indeed. Haeff gives as an estimate of the bandwidth $\Delta\omega/\omega \simeq 2{\cdot}32/(\text{gain in db})^{1/2}$.

* There are a number of papers dealing with these aspects (Pierce 1949 a, b, Pierce and Hebenstreit 1949). Other aspects of the electron-wave tube are treated by Nergaard (1948) and Rydbeck and Forsgren (1951).

12.2. *Experimental*

The first description of an electron-wave tube (Haeff 1948) mentions an electronic gain of 80 db at 3 000 Mc/s with electron streams 20 cm long of a few milliamperes. A detailed description and discussion of an electron-wave tube at 255 Mc/s is given by Hollenberg (1949), who constructed and operated such a tube and compared the results with theory (Pierce and Hebenstreit 1949). Using such a low frequency results in a high value of ω_p/ω, and electron-wave interaction effects will be very strong.

Two hollow electron beams were used (fig. 18), generated at annular cathodes, and the signal is impressed upon the inner beam by a short helix having a phase velocity equal to the inner beam velocity. After passing through the drift space a similar helix extracts the r.f. energy from the beam.

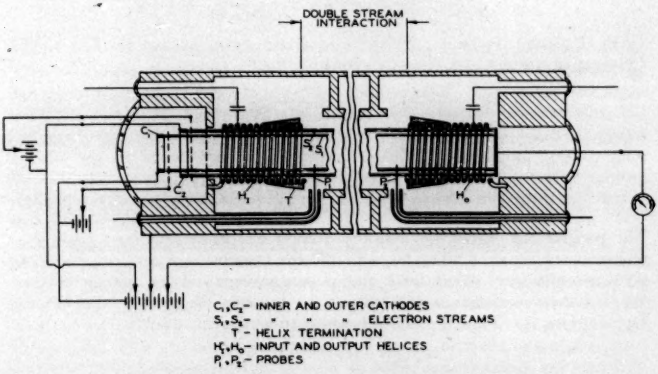

DOUBLE STREAM INTERACTION
C1, C2 – INNER AND OUTER CATHODES
S1, S2 – " " " ELECTRON STREAMS
T – HELIX TERMINATION
H1, H2 – INPUT AND OUTPUT HELICES
P1, P2 – PROBES

Fig. 18. Double-stream amplifier with helix input and output.
Acknowledgment to Bell Telephone Laboratories.

The inner beam voltage is about 50 and the outer beam travels at a lower velocity. An axial magnetic field of 700 gauss is applied to keep the beams straight. With about 1·1 mA in each beam, the outer beam at 33 volts, an overall gain of 29 db was obtained at 200 Mc/s; the maximum power output was 0·3 mW. Under similar conditions a bandwidth of 110 Mc/s, centred at 255 Mc/s, was observed.

Proof that the gain observed was due to double-beam interaction was obtained measuring the gain with each beam alone, in which case the gain was of the order of unity. Turning the other beam on, this then increased to 29 db.

A similar tube with resonator output was built, and gave, as the former tube, good agreement with the theory developed by Pierce and Hebenstreit.

Details are also given by Haeff (1949) of an electron-wave tube with *one beam only*, in which the effect of the velocity difference between two beams is achieved by the depression of the potential near the axis of the beam. This tube, called the 'whistle' tube, gave a gain comparable with that of the two-beam electron-wave tube. At present there is no rigorous theory available to describe such a tube; indeed it is claimed by Pierce (1950, chap. 16): ". . . Analysis indicates

that a multiply peaked distribution of current with velocity is necessary for the existent increasing waves, and gain in a 'single stream' of electrons is still something of a mystery".

Little experimental work has been published on the noise factor of the electron-wave tube, but one gathers that it is bad. The power output and efficiencies reported are very low, orders of magnitude below those obtained with travelling-wave tubes. No quantitative estimates are available; the cause may well be poor coupling between the electron streams and the output circuit.

§ 13. THE 'SLIPPING-STREAM TUBE'

The kind of velocity distribution which is believed to exist in the 'whistle' tube described in the previous section can be set up in a variety of ways. When it is set up in a system of crossed electric and magnetic fields it can be shown that the combined effect of the forces due to the applied d.c. fields and that due to the space charge is a kind of flow in which electrons move rectilinearly normal both to the electric and to the magnetic field directions with a velocity that is a simple function of distance. Moreover, it is found that the charge density is then uniform everywhere. Such a stream has been called a 'slipping stream'.

The simplest kind of distribution is one in which electrons move, initially, parallel to the z axis with a velocity which depends linearly on y. The stream is of infinite extent in the x direction. This system has been studied by Macfarlane and Hay (1950), who consider not only the frequency spectrum of propagation of signals in a free, non-enclosed slipping stream, but also propagation along a slipping stream coupled to a waveguide with a slow phase velocity. Let the velocity distribution be described by $u=u_0+\omega_0 y$, where $\omega_p=(e\rho_0/\epsilon m)^{1/2}$, the electron-plasma frequency. The boundaries of the stream are $y=\pm a$ (when disturbed the beam edge is presumed to oscillate about that boundary). The velocity range can thus be described in terms of a number α where

$$\alpha=(u_a-u_{-a})/2u_0=\omega_p a/u_0.$$

Macfarlane and Hay show that growing waves with rather complicated behaviour are possible over a wide range of frequencies. In particular, when $\alpha<0{\cdot}42$, the system behaves like a two-beam electron-wave tube with a cut-off frequency $\omega_1=\omega_p/\sqrt{2\alpha}$, but has only half its maximum gain, viz. $2{\cdot}1\omega_p/u_0$ db per unit length. In addition, it has a gain of $0{\cdot}53\ \omega_p/(u_0-\omega_p a)$ when $\omega\gg\omega_1$.

When $\alpha>0{\cdot}42$ the system behaves primarily like a travelling-wave tube and the rate of gain is about $0{\cdot}53\omega_p/(u_0-\omega_p a)$ db per unit length at all frequencies above ω_p.

On the same basis the gain in db per unit length of the 'ordinary' travelling-wave tube is approximately $6(\omega/\omega_p)^{1/3}\omega_p/u_0$ and that of the 'ordinary' two-beam electron-wave tube approximately $4{\cdot}3\omega_p/\bar{u}_0$, where $\bar{u}_0$ is the mean velocity of the two beams.

No experiment is known so far in which the theory mentioned above has been tested. Haeff's 'whistle' tube worked in a longitudinal magnetic field, in which case the above analysis does not apply; we have now definitely entered a region in which speculations and calculations abound, but in which experiments are scarce. It has been suggested by a number of authors that the recently discovered radio noise from the sun, from sunspots and from various locations in space may originate in some process akin to growing waves in interpenetrating electron and ion streams, in slipping streams and in a number of other configurations of moving charge distributions. Some attention has been focused

lately on the difference between a wave growing with distance and a disturbance growing with time.

In view of the unsatisfactory state of the subject at this moment, with several diametrically opposed views still unresolved, we do not propose to treat these matters at all. The interested reader is referred to the numerous important papers listed in the literature at the end of this report, and—if some modest advice may be given—it may be preferable to study the papers in the opposite order of their appearance; thus some confusion may be avoided. There is no doubt however that the subject is pregnant with new discoveries and developments, and worthy of serious attention.

§ 14. THE VELOCITY-JUMP TUBE

Recently a description of a rather intriguing device has appeared (Field, Tien and Watkins 1951) called the 'velocity-jump amplifier'.*

In this device an electron beam is velocity-modulated by a helix and allowed to drift at a potential V_1 until the resultant bunching has reverted to a pure velocity modulation. (See § 11 for a description of this mechanism.) At this point the beam is suddenly decelerated by entering a drift tube at a potential V_2 lower than V_1. It is easy to see that this must lead to an increase in the a.c. velocity, from simple energy considerations:

$$\tfrac{1}{2}\boldsymbol{m}(u_1+v_1\sin\omega t)^2=\tfrac{1}{2}\boldsymbol{m}(u_2+v_2\sin\omega t)^2-\boldsymbol{e}(V_1-V_2),$$

where u_1 and u_2 are the d.c. velocities and v_1 and v_2 the a.c. velocities in space 1 and 2 respectively. Neglecting terms in $\sin^2\omega t$ we have, therefore, $u_1v_1=u_2v_2$ or $v_2/v_1=u_1/u_2=(V_1/V_2)^{1/2}$.

The beam is allowed to drift for a quarter of a plasma wavelength at this lower potential V_2 until all the velocity modulation is converted into convection current given by $i_2=I_0v_2\omega/u_2\omega_{p_2}$, where ω_{p_2} is the electron plasma frequency in the V_2 space. At this point the potential suddenly reverts to V_1, the beam entering a tube at this potential. The a.c. current across the gap must be continuous, and thus we have $i_3=i_2$ at this point. Drifting now for a quarter of a plasma wavelength at the high potential V_2, this a.c. current will turn into a velocity modulation $v_3/u_1=i_3\omega_{p_1}/I_0\omega$.

It can thus be shown that v_3 may be considerably larger than v_1:

$$\frac{v_3}{v_1}=\left(\frac{u_1}{u_2}\right)^{3/2},\qquad \frac{v_3}{v_1}=\left(\frac{V_1}{V_2}\right)^{3/4}.$$

This, then, is the 'voltage' gain per stage (a stage consisting of a high-potential drift tube and a low-potential drift tube, each an odd number of quarter wavelengths of the respective plasma frequency long). The power gain per stage is the square of this, namely $(V_1/V_2)^{3/2}$.

The remarkable thing about this tube is the fact that amplification is obtained without double beams or slipping streams, and without any material structure supporting a slow wave.

A question may be asked: What then is the mechanism by which d.c. energy is converted into r.f. energy? This question has been dealt with in discussing the klystron and the travelling-wave tube, and it has been mentioned that the electron-wave tube can be shown to have a mechanism essentially similar to that

* The author is indebted to Professor L. M. Field, Electronics Research Laboratory, Stanford University, for information on this tube.

of the travelling-wave tube. With the velocity-wave jump tube the mechanism is not so easily seen, but is believed to be somewhat as follows: with a space-charge wave there are associated four a.c. quantities, namely velocity, convection current, charge and voltage. The last quantity provides the clue. It can be shown that the a.c. potential on either side of the potential jump from V_2 to V_1 is *not* continuous, though the convection current there is continuous. Thus we have a discontinuity of a.c. potential at this jump (in practice of finite extent) which is just the kind of a.c. field found in the catcher of the klystron. The a.c. convection current is passing through this a.c. field in such a phase that more electrons are retarded than accelerated. Thus energy is taken from the d.c. energy of the beam and converted into a.c. energy in the beam. This secondary a.c. field will itself velocity-modulate the electron beam, and a complete theory of the velocity-jump tube must include effects like these, apart from considering the effect of the geometry of beam, drift tubes and gaps.

In an experimental single-stage amplifier of this type, a net gain of 24 db was observed with V_1 1 900 v, V_2 65 v, I_0 0·6 mA, frequency 3 000 Mc/s. The bandwidth of this type of tube, though very large, is limited at low frequencies because the effective plasma frequency is then reduced on account of the finite beam cross section. At the high frequencies gain falls off because it becomes difficult to excite the first-order plasma waves. Actually, the means of coupling into and out of the beam, such as helices or resonators, have usually far narrower bandwidths than the electronic amplification process itself.

The principle of influencing the amplitude of electron-plasma waves by means of velocity jumps can, of course, also be employed in reverse, that is to say, to *decrease* the amplitude of a wave. Now we know that shot noise and thermal noise in an electron beam exist in the form of electron plasma waves (Cutler and Quate 1950), and thus should be amenable to reduction by the principle of velocity jumps.

Field and Watkins (private communication) have indeed applied this principle to low-noise travelling-wave tubes and have obtained an improvement of the noise factor of 5·7 db with a single-stage velocity-jump de-amplifier placed before the input section of a travelling-wave tube. The theoretical improvement should have been 8·2 db, and the discrepancy is ascribed to partition noise and non-uniform cathode emission. The operating characteristics of this travelling-wave tube were: V_1 190 v, V_2 2 600 v, I_0 175 mA, noise factor 11·3 db.

Unfortunately the improvement in noise factor obtainable in this way is not unlimited. It has been shown (Robinson 1952) that the noise content of an electron beam consists of two uncorrelated components one-quarter of a plasma wavelength out of phase, and the mechanism discussed above, while de-amplifying one component, will amplify the other.

§15. OTHER FORMS OF TRAVELLING-WAVE TUBES

15.1. *Transverse Current Travelling-Wave Tube*

This device is a travelling-wave tube employing a flat sheet electron beam close to the flat surface of a helix wound on a former of rectangular cross section (Field 1950). The windings are skew or slanting, as shown in fig. 19. The direction of propagation along this helix is at right angles to the direction of motion of the electron beam; because of the skewness of the windings there is, however, a

component of electric field in the direction of the beam, and the electron velocity is adjusted to a value where interaction with the wave is obtained through this component.*

Comparing this tube with the conventional travelling-wave tube several differences are found. At low current the gain is not proportional to the cube root of the beam current, but is proportional to the square of the current. At high current the gain becomes independent of current. Similarly, the power output is not proportional to the four-thirds power of beam current, but reaches a maximum level which is, moreover, independent of beam current at high currents, and also independent of input power provided there is sufficient input power.

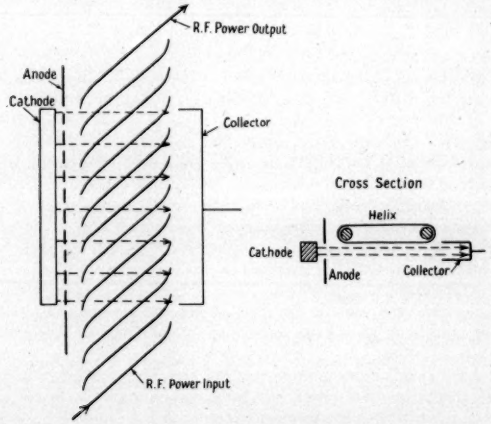


Fig. 19. Sketch of transverse-current travelling-wave tube.

The tube is thus capable of functioning as a high-level limiter over a wide band of frequencies. Qualitatively this can be explained by considering the current as made up of a series of small current elements, each of which may reach saturation in the usual fashion, space-charge repulsion or 'debunching' limiting the maximum a.c. current that can be excited in each element. At some point along the length of the tube it is to be expected that a current element will gain as much energy in the debunching part of its travel as it imparts to the helix in the bunching part of its travel. Beyond this point it is reasonable to expect all subsequent current elements to behave similarly. Thus an equilibrium power level is reached, and no increase or decrease beyond this point occurs.

Experimental values for a typical transverse current travelling-wave tube are as follows: beam voltage 96 v, beam current 47 mA, mid-frequency 161 Mc/s, net gain 31 db, maximum power output 271 mW, helix cross section 0·62 cm × 6·10 cm, slant angle 35°, bandwidth of match to helix (for an input standing wave ratio less than 3·5) 120–500 Mc/s.

* The writer is indebted to Professor L. M. Field for all the information on this tube.

15.2. *The 'Easitron'*

For all the information on this new conception of a microwave amplifier the writer is indebted to Dr. J. R. Pierce, who kindly put at his disposal a Bell Telephone Laboratories memorandum on this subject by L. R. Walker.

The material structure, with which the electron beam is made to interact, consists of an array of closely spaced thin wires half a wavelength long, connecting two infinitely large conducting planes, as in fig. 20. The propagation characteristics of this kind of structure have been discussed by Pierce (1950, chap. 5).

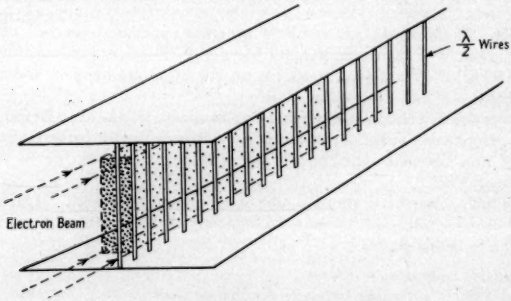


Fig. 20. Sketch of the 'Easitron'. The electron beam is shot along the wires which connect two plates half a wavelength apart.

The structure cannot, by itself, propagate a travelling wave, but if an electron beam is projected alongside the wires (normal to the direction of individual wires), it has been shown that a space-charge wave in the beam will be strongly amplified, *independently* of the beam velocity. The electron beam supplies the coupling between wires needed to obtain propagation and also d.c. energy to make the wave grow. The wire structure supplies an 'inductive component' of space charge which more than compensates the normal capacitive component of space charge, so that, in effect, there is now a tendency for equal charges to *attract* each other, instead of repelling each other. (We are, of course, only considering electron velocities well below **c**.) Thus any bunching existing in the beam before reaching the parallel wire array will tend to become *denser* on interacting with the structure; the amplitude of any signal in the beam will increase and the tube will amplify.

Since there is no synchronism between an unperturbed wave and an electron beam, the 'Easitron' will operate at any beam voltage; however, it will be sharply 'tuned' to one frequency, determined by the distance between the two conducting planes.

Expressions for the gain of such a tube are given in the above-mentioned memorandum, and it appears that the gain may be of the same order of magnitude as in a comparable electron-wave tube. Attenuation in the material structure appears to be unimportant, and feedback is, of course, absent.

No experimental details are available of this tube, though the writer ventures to surmise that it must be a tube of very uncritical operation, and very *easy* to make to work—hence the name 'Easitron'.

§ 16. APPLICATIONS OF THE TRAVELLING-WAVE TUBE

The most important application of the travelling-wave tube is in the field of communications, and more specifically in the field of microwave communications, which has been pointed out by J. R. Pierce—and the introduction to his book on travelling-wave tubes explains the position very clearly. The transmission of information needs a band of frequencies, to a first order proportional to the amount of information. More sophisticated methods of transmission, such as frequency modulation or pulse-code modulation, need even more bandwidth; in return for this they are less susceptible to inference by noise.

At present the potential bandwidth of the travelling-wave tube is by no means fully used. If, however, the trend of progress continues, the full bandwidth available, namely of the order of 1 000 Mc/s at 4 000 Mc/s mid-band frequency, will soon be filled with television, telephone and facsimile channels, and engineers and scientists will look around for more.

As far as the writer is aware, the first communication link to use travelling-wave tubes in practice anywhere in the world is the British Broadcasting Corporation's television link between Manchester and Edinburgh, due to be in operation some time in 1952.

Other links, elsewhere, are no doubt being planned and built. The following are some other applications of the travelling-wave tube which are known to have been made in actual practice:

(*a*) *As tunable laboratory oscillator.*

(Field 1949) By using external feedback incorporating a tunable filter, a travelling-wave tube was made to oscillate between 7 500 and 11 750 Mc/s. Tuning was continuous, and output of second harmonic power of several milliwatts was obtained between 15 000 and 23 500 Mc/s.

(van Iperen 1950) This tube was developed mainly to check a theory of oscillation in travelling-wave tubes developed earlier (van Iperen 1949). However, a tube oscillating strongly in the 20 cm region and tunable by varying the beam voltage should be useful as a source of r.f. power. It is of interest to note that gas focusing was employed to hold the beam together.

(*b*) *As generator of very short pulses.*

(Klute 1951) In this application use is made of the fact that gain in the travelling-wave tube is a rapidly varying function of beam voltage. A c.w. signal at 4 000 Mc/s is injected into a travelling-wave tube whose beam voltage is varied sinusoidally at 18·432 Mc/s with a peak amplitude of 300 volts. This produces pulses, 0·005 microsecond long, repeated every 0·054 microsecond with a peak power presumably determined by that of the travelling-wave tube, namely of the order of watts. It is stated that pulses as short as 0·001 microsecond have been generated.

(*c*) *Laboratory amplifier for wide-range tunable oscillators.*

(Radcliffe 1951) Apparatus used to test microwave circuits includes a triode oscillator whose frequency is swept mechanically over a range of 500 Mc/s around 4 000 Mc/s. The output from the triode is low and varies over the band. Therefore it is fed into a travelling-wave tube where it is amplified up to the level of 100 milliwatts. A portion of this power is rectified and fed back to control the power gain of the travelling-wave tube in such a way as to produce an output constant to within $\pm 0{\cdot}02$ db over the whole band.

(*d*) *Wide-band high-frequency oscilloscopes.*

(Pierce 1949) A very well focused electron beam is projected through a delay line built in the form of a periodic filter network. A signal of 0·037 volt into the input impedance of 75 ohms deflects the luminous spot by one spot diameter. The screen is viewed through a 60× microscope. Frequencies from zero up to 500 Mc/s are faithfully reproduced.

(Owaki, Terehata, Hada and Nakamura 1950) In this tube Lecher-lines folded zigzag form both *x* and *y* deflecting plates and are intended to slow down the input signals to the same velocity as the electron beam. No experimental figures of bandwidth and sensitivity are given.

REFERENCES

Introduction and Historical Background

BOOT, H. A. H., and RANDALL, J. T., 1946, *J. Instn. Elect. Engrs.*, **93,** Part IIIA, 928.

COLLINS, G. B., 1948, *Microwave Magnetrons, Radiation Lab. Series* (New York: McGraw-Hill).

FISK, J. B., HAGSTRUM, H. D., and HARTMAN, P. L., 1946, *Bell Syst. Tech. J.*, **25,** 167.

FOSTER, J., 1946, *J. Instn. Elect. Engrs.*, **93,** Pt. IIIA, 192, 868.

FRY, D. W., and WALKINSHAW, W., 1949, *Rep. Prog. Phys.*, **12,** 102 (London: Physical Society).

HAHN, W. C., and METCALF, G. F., 1939, *Proc. Inst. Radio Engrs., N.Y.*, **27,** 106.

HAMILTON, D. R., KNIPP, J. K., and KUPER, J. B. H., 1948, *Klystrons and Microwave Triodes, Radiation Lab. Series* (New York: McGraw-Hill).

HANSEN, W. W., 1945, *J. Appl. Phys.*, **16,** 528.

HEIL, A., and HEIL, O., 1935, *Z. Phys.*, **95,** 752.

KOMPFNER, R., 1947, *J. Brit. Instn. Radio Engrs.*, **7,** 117.

NEHER, H. V., 1943, *M.I.T. Radiation Lab. Report* No. 306; 1945, *Ibid.*, No. 822.

SLOAN, D. H., and LAWRENCE, E. O., 1931, *Phys. Rev.*, **38,** 2021.

VARIAN, R. H., and VARIAN, S. F., 1939, *J. Appl. Phys.*, **10,** 140.

WHEELER, H. A., 1939, *Proc. Inst. Radio Engrs., N.Y.*, **27,** 429.

WIDERÖE, R., 1928, *Arch. Elektrotech.*, **21,** 387.

Travelling-Wave Tubes, Theoretical and Experimental

AGDUR, B. N., and ÅSDAL, C. G. L., 1951, *Trans. Chalmers Univ. Tech.* No. 106.

BERNIER, J., 1947 a, *Ann. Radioélectricité*, **2,** 87; 1947 b, *Onde Elect.*, **27,** 231.

BERTEROTTIERE, R., and CONVERT, G., 1950, *Ann. Radioélectricité*, **5,** 168.

BLANC-LAPIERRE, A., KUHNER, M., LAPOSTOLLE, P., JESSEL, M., and WALLAUSCHEK, R., 1948, *Ann. Télécommun.*, **3,** 257.

BLANC-LAPIERRE, A., and LAPOSTOLLE, P., 1946, *Ann. Télécommun.*, **1,** 283.

BLANC-LAPIERRE, A., LAPOSTOLLE, P., VOGE, J. P., and WALLAUSCHEK, R., 1947, *Onde Elect.*, **27,** 194.

BRILLOUIN, L., 1948, *Phys. Rev.*, **74,** 90; 1949, *J. Appl. Phys.*, **20,** 1196.

BRÜCK, L., 1949, *Ann. Radioélectricité*, **4,** 222; 1950, *Arch. Elektrotech.*, **39,** 633.

BRYANT, J. H., 1950, *Elect. Commun.*, **27,** 277.

CHU, L. J., and JACKSON, J. D., 1948, *Proc. Inst. Radio Engrs., N.Y.*, **36,** 853.

CUTLER, C. C., 1951, *Proc. Inst. Radio Engrs., N.Y.*, **39,** 914.

DOEHLER, O., and KLEEN, W., 1948 a, *Ann. Radioélectricité*, **3,** 124; 1948 b, *Ibid.*, **3,** 184; 1949 a, *Ibid.*, **4,** 76; 1949 b, *Ibid.*, **4,** 216; 1949 c, *Arch. elekt. Übertragung*, **3,** 54, Part 1; 1950, *Ibid.*, **4,** 207, Part 2.

FLETCHER, R. C., 1950, *Proc. Inst. Radio Engrs., N.Y.*, **38,** 413.

FRIEDMAN, B., 1951, *J. Appl. Phys.*, **22,** 443.

GOUDET, G., 1948, *Ann. Télécommun.*, **3,** 445.

HATTON, J., 1946, *Microwave Tube Conference, Yale.*

KOMPFNER, R., 1946, *Wireless World*, **52,** 369; 1947 a, *Proc. Inst. Radio Engrs., N.Y.*, **35,** 124; 1947 b, *Wireless Engr.*, **24,** 255; 1950, *J. Brit. Instn. Radio Engrs.*, **10,** 283.

LAPLUME, J., 1949, *Onde Elect.*, **29,** 66.

LAPOSTOLLE, P., 1948, *Ann. Télécommun.*, **3**, 57, 85.
LITTLE, J. B., 1951, *Bell Lab. Record*, **29**, 14.
LOSHAKOV, L. N., 1949, *Zh. Tech. Fis.*, **19**, 578.
MATHEWS, W. E., 1951 a, *J. Appl. Phys.*, **22**, 311; 1951 b, *Proc. Inst. Radio Engrs., N.Y.*, **39**, 1044.
PARZEN, P., and GOLDSTEIN, L., 1951, *J. Appl. Phys.*, **22**, 398.
PIERCE, J. R., 1946, *Bell Lab. Record*, **24**, 439; 1947, *Proc. Inst. Radio Engrs., N.Y.*, **35**, 111; 1948 a, *Ibid.*, **36**, 993; 1948 b, *Bell Syst. Tech. J.*, **27**, 732; 1950, *Traveling-Wave Tubes* (New York : Van Nostrand).
PIERCE, J. R., and FIELD, L. M., 1947, *Proc. Inst. Radio Engrs., N.Y.*, **35**, 108.
ROBINSON, F. N. H., 1951, *Wireless Engr.*, **28**, 110.
ROGERS, D. C., 1949, *Elect. Commun.*, **26**, 144.
RYDBECK, O. E. H., 1948, *Ericsson Technics*, **46**, 3.
SCHNITGER, H., and WEBER, D., 1949, *Frequenz*, **3**, 189.
SHULMAN, C., and HEAGY, M. S., 1947, *R.C.A. Rev.*, **8**, 585.
SLATER, J. C., 1950, *Microwave Electronics* (New York : Van Nostrand).
TOMNER, J. S. A., 1948, *Elect. Engng.*, **1**, No. 6, 1.
VOGE, J. P., 1946, *C. R. Acad. Sci., Paris*, **223**, 1117.
WARNECKE, R., DOEHLER, O., and KLEEN, W., 1951, *Wireless Engr.*, **28**, 167.
WEBBER, S. E., 1950, *Electronics*, **23**, 100.

The Helix as a Guide of Slow Waves

CUTLER, C. C., 1948, *Proc. Inst. Radio Engrs., N.Y.*, **36**, 230.
HSU, C. P., *Cruft Laboratory Technical Report*, No. 95.
JOHNSEN, K., and DAHL, H., 1951, *Chr. Michelsens Institutt Beretningar*, **14**, 4.
LENZ, W., 1914, *Ann. Phys., Lpz.*, **43**, 749.
LOSHAKOV, L. N., and O'DEROGGE, —, 1948, *Radioteknika (Moscow)*, **3**, 11.
LUND, C. O., 1950, *R.C.A. Rev.*, **11**, 133.
OLLENDORF, F., 1925, *Die Grundlagen der Hochfrequenztechnik* (Berlin : Springer).
PHILLIPS, R. S., 1950, *Quart. Appl. Math.*, **8**, 229.
PHILLIPS, R. S., and MALIN, H., *New York University Mathematics Research Group*, Report No. 170–3.
POCKLINGTON, H. C., 1897, *Proc. Camb. Phil. Soc.*, **9**, 324.
ROUBINE, E., 1947, *C. R. Acad. Sci., Paris*, **224**, 1101, 1149, *Onde Elect.*, **27**, 203; 1951, *C. R. Acad. Sci. Paris*, **232**, 1748.
SOLLFREY, W., 1951, *J. Appl. Phys.*, **22**, 905.
SOMMERFELD, A., 1907, *Ann. Phys., Lpz.*, **24** (4), 609.

Noise in Travelling-Wave Tubes

LLEWELLYN, F. B., and PETERSON, L. C., 1944, *Proc. Inst. Radio Engrs., N.Y.*, **32**, 144.
MACDONALD, D. K. C., 1949 a, *Rep. Prog. Phys.*, **12**, 36 (London : Physical Society); 1949 b, *Phil. Mag.*, **40**, 561; 1950, *Ibid.*, **41**, 863.
RACK, A. J., 1938, *Bell Syst. Tech. J.*, **17**, 592.
ROBINSON, F. N. H., 1952, *Phil. Mag.*, **43**, 51.
ROBINSON, F. N. H., and KOMPFNER, R., 1951, *Proc. Inst. Radio Engrs., N.Y.*, **39**, 918.

Electron Optics in the Travelling-Wave Tube

BRILLOUIN, L., 1945, *Phys. Rev.*, **67**, 260.
FIELD, L. M., 1946, *Rev. Mod. Phys.*, **18**, 353.
PIERCE, J. R., 1940, *J. Appl. Phys.*, **11**, 548; 1949, *Theory and Design of Electron Beams* (New York : Van Nostrand).
SAMUEL, A. L., 1949, *Proc. Inst. Radio Engrs., N.Y.*, **37**, 1252.
WANG, C. C., 1950, *Proc. Inst. Radio Engrs., N.Y.*, **38**, 135.

Structures other than a Helix Propagating Slow Waves, and Tubes Employing Them

BANOS, A., SAXON, D. S., and GRUEN, H., 1951, *J. Appl. Phys.*, **22**, 117.
BRILLOUIN, L., 1948, *J. Appl. Phys.*, **19**, 1023.
BRUCK, G. G., and WICHER, E. R., 1947, *J. Appl. Phys.*, **18**, 766.
CHU, L. J., and HANSEN, W. W., 1947, *J. Appl. Phys.*, **18**, 996.

DEWEY, G. C., 1948, *Proceedings of National Electronics Conference (Chicago)*, **4**, 253.
DEWEY, G. C., PARZEN, P., and MARCHESE, T. J., 1951, *Proc. Inst. Radio Engrs., N.Y.*, **39**, 153.
FIELD, L. M., 1949, *Proc. Inst. Radio Engrs., N.Y.*, **37**, 34.
FRANKEL, S., 1947, *J. Appl. Phys.*, **18**, 650.
KETTEL, E., 1949, *Frequenz*, **3**, 73.
MILLMAN, S., 1951, *Proc. Inst. Radio Engrs., N.Y.*, **39**, 1035.
PIERCE, J. R., 1949, *Proc. Inst. Radio Engrs., N.Y.*, **37**, 510.
PIERCE, J. R., and WAX, N., 1949, *Proc. Inst. Radio Engrs., N.Y.*, **37**, 622.
SLATER, J. C., 1946, *Rev. Mod. Phys.*, **18**, 441.

Various Forms of Tubes based on the Travelling-Wave Principle: Latest Developments

BROSSART, J., and DOEHLER, O., 1948, *Ann. Radioélectricité*, **3**, 328.
FIELD, L. M., 1950, *Electronics*, **23**, 100.
FIELD, L. M., PING KING TIEN, and WATKINS, D. A., 1951, *Proc. Inst. Radio Engrs., N.Y.*, **39**, 194.
KLINGER, H. H., 1949, *Elektrotechnik*, **3**, 90; 1951, *Arch. elekt. Übertragung*, **5**, 167.
MOURIER, G., 1950, *Ann. Radioélectricité*, **5**, 206.
WARNECKE, R., and DOEHLER, O., 1950, *C. R. Acad. Sci., Paris*, **231**, 1132.
WARNECKE, R., and GUENARD, P., 1948, *Ann. Radioélectricité*, **3**, 259.
WARNECKE, R. R., KLEEN, W., LERBS, A., DOEHLER, O., and HUBER, H., *Proc. Inst. Radio Engrs., N.Y.*, **38**, 486.

Plasma Waves and Electron Wave Tubes

AGDUR, B. N., 1951, *Trans. Chalmers Univ. Technol.* No. 105.
BAILEY, V. A., 1948, *Austral. J. Sci. Res.*, **1**, 351; 1949, *J. Roy. Soc. N.S. Wales*, **82**, 107; 1950, *Phys. Rev.*, **78**, 428; 1951, *Ibid.*, **83**, 439.
BOHM, D., and GROSS, E. P., 1949, *Phys. Rev.*, **75**, 1851.
BRILLOUIN, L., 1946, *Phys. Rev.*, **70**, 187.
FEINSTEIN, J., and SEN, H. K., 1951, *Phys. Rev.*, **83**, 405.
GUENARD, P., BERTEROTTIERE, R., and DOEHLER, O., 1949, *Ann. Radioélectricité*, **4**, 171.
HAEFF, A. V., 1948, *Phys. Rev.*, **74**, 1532; 1949, *Proc. Inst. Radio Engrs., N.Y.*, **37**, 4.
HOLLENBERG, A. V., 1949 a, *Bell Lab. Record*, **27**, 290; 1949 b, *Bell Syst. Tech. J.*, **28**, 52.
LABUS, J., 1950, *Arch. elekt. Übertragung*, **4**, 353.
MACFARLANE, G. G., and HAY, H. G., 1950, *Proc. Phys. Soc.* B, **63**, 409.
MACFARLANE, G. G., and WOODWARD, A. M., 1950, *Proc. Instn. Elect. Engrs.*, **97**, Pt. III, 322.
MALMFORS, K. G., 1950, *Ark. Fys.*, **1**, 30, 569.
NERGAARD, L. S., 1948, *R.C.A. Rev.*, **9**, 585.
PIERCE, J. R., 1948, *J. Appl. Phys.*, **19**, 231; 1949 a, *Ibid.*, **20**, 1060; 1949 b, *Proc. Inst. Radio Engrs., N.Y.*, **37**, 980; 1950, *J. Appl. Phys.*, **21**, 1063; 1951 a, *Physics Today*, **4**, 4; 1951 b, *Bell Syst. Tech. J.*, **30**, 626.
PIERCE, J. R., and HEBENSTREIT, W. B., 1949, *Bell Syst. Tech. J.*, **28**, 33.
RAMO, S., 1939 a, *Phys. Rev.*, **56**, 276; 1939 b, *Proc. Inst. Radio Engrs., N.Y.*, **27**, 757.
ROBERTS, J. A., 1949, *Phys. Rev.*, **76**, 340.
RYDBECK, O. E. H., and FORSGREN, S. K. H., 1951, *Trans. Chalmers Univ. Tech.*, No. 102.
SCHUMANN, W. O., 1950, *Z. angew. Phys.*, **2**, 393.
TONKS, L., and LANGMUIR, I., 1929, *Phys. Rev.*, **33**, 195.
TWISS, R. Q., 1951, *Proc. Phys. Soc.* B, **64**, 654.
WALKER, L. R., 1949, *Phys. Rev.*, **76**, 1721.

Applications of Travelling-Wave Tubes

DOEHLER, O., KLEEN, W., and PALLNEL, P., *Ann. Radioélectricité*, **4**, 68.
VAN IPEREN, B. B., 1949, *Philips Res. Rep.*, **4**, 20; 1950, *Philips Tech. Rev.*, **11**, 221.
KLUTE, W. A., 1951, *Bell Lab. Record*, **29**, 216.
OWAKI, K., TERAHATA, S., HADA, T., and NAKAMURA, T., 1950, *Proc. Inst. Radio Engrs., N.Y.*, **38**, 1172.
PIERCE, J. R., 1949, *Electronics*, **22**, 97.
RADCLIFFE, F. E., 1951, *Electronics*, **24**, 110

REPORTS ON PROGRESS IN PHYSICS

CUMULATIVE SUBJECT INDEX, VOLS. 1–15

Absolute electrical measurements 5, 302
Absorption of gases by solids 4, 27
Absorption of light in interstellar space 7, 150
Absorption spectroscopy, ultra-violet, recent advances 14, 1
Accelerators, cyclic 13, 295
Accelerators, linear 12, 102
Acoustics, musical 7, 27
Angular correlations in nuclear reactions 14, 196
Anomalous fine structure of hydrogen and singly ionized helium 14, 19
Astronomy, recent progress in 5, 82
Astrophysics, instrumental technique 8, 186
Atomic physics 2, 74; 3, 66; 4, 173; 5, 348
Atomic spectra, multipole radiation in 12, 233
Atomic structures, calculation of 11, 113
Aurora, spectrum of the airglow 14, 121

Band structure of metals 15, 173
Bohr theory of nuclear reactions 7, 87
Burgers' theory of turbulence 1, 43

Capacitance, measurement of 6, 403
Centrifuging, high-speed 8, 31
Charge of the electron 2, 247; 4, 212
Clouds, microphysics of 14, 147
Collisions between atoms and molecules at ordinary temperatures 12, 248
Colloidal systems, aqueous, phase separation in 10, 129
Combustion, applications of spectroscopy to 8, 50
Comets and meteor-streams 11, 402
Conductors, galvanomagnetic effects in 15, 249
Conservation of energy and momentum, elementary processes 3, 89
Cosmic rays 10, 1
Counter tubes, recent research on 9, 137
Counting, electrical methods 3, 239
Crystal dynamics and X-ray scattering 9, 294
Cyclic accelerators 13, 295
Cyclotron and some of its applications 6, 125

Design, mechanical, of physical instruments 10, 272
Development of the teaching of experimental physics in British universities . 7, 355
Diamagnetic and paramagnetic anisotropy of crystals 4, 368
Dielectric breakdown in solids 6, 411
Diffraction theory of aberrations 14, 95
Discharge phenomena in gases 8, 338
Dispersion in the far infra-red 8, 200

Elasticity 3, 9
Electric and magnetic measurements 1, 321
Electric wave filters 5, 334
Electrical and magnetic measurements 2, 221
Electrical discharges in gases and their applications 7, 230
Electrical methods of counting 3, 239
Electrical phenomena at extremely low temperatures 1, 198

Electricity and magnetism, experimental 3, 206; 4, 218
Electrolytes and electrolysis 4, 251
Electron microscopes of high magnification 7, 107
Electron microscopy 10, 204
Electron tubes 2, 258
Electrostatic generators for acceleration of charged particles 11, 1
Elementary particles, theory 10, 253
Equation of state of gases and molecular distribution 12, 305
Equations of state 7, 195
Evaporation in nature 11, 366
Eye movements, investigation 13, 1

Ferrites 15, 142
Ferromagnetism 11, 43
Ferromagnetism : magnetization curves 13, 83
Field emission of electrons 9, 177
Fluid motion 3, 26
Fracture and strength of solids 12, 185
Friction between solid bodies 8, 71

Galvanomagnetic effects in conductors 15, 249
Gamma radiations emitted in nuclear processes 7, 66
Gassiot Committee of the Royal Society, reports 9, 1
Geiger counter, mechanism of 5, 390
General physics (classical) 2, 1; 3, 1; 5, 1
Geophysical prospecting 2, 97
Glass technology, physics in 11, 342
Glass, viscosity and related properties in 12, 133

Hearing, defective, instrumental aids 5, 150
Heat 1, 161; 2, 188; 3, 143; 4, 76; 5, 164; 6, 297
High energy nucleon–proton and proton–proton scattering data 15, 68
Hydrogen bond 15, 1

Impedance networks 6, 389
Infra-red, far, dispersion in 8, 200
Infra-red region of the spectrum, developments 11, 144
Infra-red solar spectroscopy, recent advances 13, 24
Infra-red spectra of polyatomic molecules 7, 41
Insoluble monolayers at the air–water and oil–water interfaces 9, 158
Instrumental aids for defective hearing 5, 150
Instrumental technique in astrophysics 8, 186
Interactions of gases with metals and crystalline solids 7, 303
Ionospheric temperatures, critical survey 14, 316
Isotopes, separation of 6, 48

Kinetic theory of the elasticity of rubber 6, 330

Latent image formation in photographic silver halide gelatine emulsions . . 11, 248
Lens systems, new 7, 130
Lightning discharge 10, 314
Linear accelerators 12, 102
Liquid helium, properties of 6, 280
Liquid state, theories of 5, 46
Low temperature physics 12, 270
Luminescence of solids 6, 174

Magnetism 3, 175
Mass spectrometry 12, 1; 15, 24
Measurement of noise 3, 130
Measurement of time 4, 1
Mechanism of the Geiger counter 5, 390
Medicine, nuclear physics in 14, 366
Medicine, use of x-rays and γ-rays 5, 284
Mesons 6, 78; 13, 350
Metals, band structure 15, 173
Metals, theory, and superconductivity 3, 262
Meteor ionization theory 11, 444
Meteoric ionization and ionospheric abnormalities 11, 415
Meteorology 5, 100
Meteors, comets and meteoric ionization 11, 389
Microphysics of clouds 14, 147
M.I.T. wavelength project 8, 212
Molecular distribution and equation of state of gases 12, 305
Molecular electronic spectra, dispersion and polarization 8, 231
Molecular structure, theory of 6, 212
Multipole radiation in atomic spectra 12, 233

Neutrino and the recoil of nuclei in beta disintegrations 11, 32
Noise, measurement of 3, 130
Nuclear effects in atomic spectra 14, 288
Nuclear interactions of cosmic rays 14, 227
Nuclear levels, symmetry properties of 8, 274
Nuclear paramagnetism 12, 22
Nuclear physics in medicine, critical survey 14, 366
Nuclear reactions, angular correlations 14, 196
Nucleon–proton scattering, high energy 15, 68

Oil, search for, and physics 9, 198
Optical interferometry, new techniques 14, 64
Optics 1, 131; 2, 356; 3, 290; 5, 182
Oxidation of metals, theory 12, 163
Oxide-coated cathodes, mechanism of thermionic emission 11, 298

Paintings, physics of 9, 354
Paramagnetic relaxation effects 13, 276
Phase separation in aqueous colloidal systems 10, 129
Phosphors and phosphorescence 12, 34
Photoelectricity 3, 272
Photometry 8, 318
Physical constants, general 8, 90
Physical-optical instruments and materials 4, 283
Physics of the atom 1, 269
Physics, general 2, 1; 3, 1; 5, 1
Physics of paintings 9, 334
Physics and the search for oil 9, 198
Physics and textiles 10, 141
Physics and textiles, processing 10, 172
Planetary and satellite atmospheres 13, 247
Plastics in industrial physics 5, 64
Properties of matter, ultrasonics research 11, 205
Proton–proton scattering, high energy 15, 68

Quantum mechanics, difficulties and developments **5**, 402
Quantum theory **2**, 27
Quantum theory, new, beginnings **4**, 134
Quantum and wave mechanics **1**, 1

Radioactivity branching **11**, 19
Radioactivity, induced **6**, 16
Radio astronomy **13**, 184
Radio-frequency spectroscopy of gases **11**, 178
Range–energy relations **15**, 49
Reactions in solids **6**, 186
Refrigeration **4**, 124
Rubber, kinetic theory of the elasticity of **6**, 330
Rubber, structure and elasticity of **9**, 113

Scattering data, high-energy neutron–proton and proton–proton . . . **15**, 68
Seismology **10**, 52
Shock waves and the propagation of finite pulses in fluids **13**, 46
Short-wave radiations, production and measurement of **6**, 95
Slow neutrons **4**, 198
Slow neutron absorption cross sections of the elements **12**, 291
Solar physics **7**, 160; **9**, 101
Solids, strength and fracture **12**, 185
Sound **1**, 63; **2**, 166; **3**, 106; **4**, 53; **5**, 121; **6**, 238; **7**, 1; **9**, 228; **10**, 120
Specific heat of solids, theory of **8**, 11
Spectrometry, mass **12**, 1; **15**, 24
Spectrophotometry, photo-electric, and its use in industry **7**, 36
Spectroscopic observations on comets, meteors and meteorites **11**, 410
Spectroscopy **1**, 102; **2**, 329; **3**, 369; **4**, 295; **5**, 210; **6**, 137
Spectroscopy, applications to combustion **8**, 50
Spectroscopy applied to molecular physics **10**, 378
Spectroscopy, radio-frequency, of gases **11**, 178
Spectrum of the airglow and the aurora **14**, 121
Spiral nebulae and the expansion of the universe **1**, 24
Spontaneous fluctuations **12**, 56
Stellar interiors and stellar evolution, physics of **6**, 1
Superconductivity **6**, 335; **10**, 358
Superconductivity and the theory of metals **3**, 262
Supersonics in relation to molecular constitution **4**, 67
Surface tension **1**, 30; **3**, 1; **4**, 40; **7**, 180

Teaching of physics in schools **5**, 422
Teaching of physics in technical institutions **6**, 431
Teaching of theoretical physics in universities **8**, 1
Television developments **8**, 135
Textiles and physics **10**, 141
Textiles, processing, applications of physics **10**, 172
Thermodynamics **4**, 103
Time, measurement of **4**, 1
Travelling-wave tubes **15**, 275
Turbulence, Burgers' theory **1**, 43
Turbulent motion **15**, 101

Upper atmosphere **3**, 42
Upper atmospheric ionization, radio exploration **2**, 129
Ultra-short electric waves on wires **6**, 378
Ultrasonics research and properties of matter **11**, 205

Viscosity **3,** 19
Viscosity and the nature of substances of high molecular weight in solution . **7,** 329
Viscosity and related properties in glass **12,** 133

War, lessons of, for science **10,** 418

X-ray analysis and application of Fourier series method to molecular structures . **4,** 332
X-ray diffraction techniques in the industrial laboratory **10,** 83
X-ray scattering in relation to crystal dynamics **9,** 256
X-ray spectroscopy of the solid state **5,** 257
X-rays **1,** 228; **2,** 301; **3,** 338
X-rays and crystals **6,** 345
X-rays and γ-rays in medicine **5,** 284
X-rays, high-voltage, recent nuclear experiments **12,** 82

CUMULATIVE AUTHOR INDEX, VOLS. 1–15

Alexander, A. E. : Insoluble monolayers at the air–water and oil–water interfaces 9, 158
Anderson, J. S. (co-author) ; Optics 1, 131
Andrade, E. N. da C. : Physics of the atom 1, 269
Appleton, E. V. : Radio exploration of upper-atmospheric ionization . . 2, 129
Astbury, N. F. : Electric wave filters 5, 334
Astbury, N. F. : The measurement of capacitance 6, 403
Awbery, J. H. : Heat 1, 161; 2, 188
Awbery, J. H. : Thermodynamics 4, 103

Backhurst, I. (co-author) : X-rays 1, 228; 2, 301; 3, 338
Bamford, C. H. : Photo-chemical processes in an oxygen–nitrogen atmosphere . 9, 75
Barr, G. (co-author) : General physics 3, 1
Barton, A. W. : The teaching of physics in schools 5, 422
Batchelor, G. K. : Turbulent motion 15, 101
Bates, D. R., *see* Massey, H. S. W.
Bates, L. F. : Magnetism 3, 175
Beams, J. W. : High-speed centrifuging 8, 31
Beattie, A., and Stockmayer, W. H. : Equations of state 7, 195
Bell, G. E. (co-author) : X-rays 1, 228; 2, 301; 3, 338
Bell, G. E. (co-author), *see* Binks, W.
Bell, G. E., *see* Kaye, G. W. C.
Berg, W. F. : Latent image formation in photographic silver halide gelatine emulsions 11, 248
Bernal, J. D. : Lessons of the war for science 10, 418
Bethe, H. A., and Marshak, R. E. : The physics of stellar interiors and stellar evolution 6, 1
Bhabha, H. J. : The theory of the elementary particles 10, 253
Binks, W. (co-author) : X-rays 2, 301; 3, 338
Binks, W., and Bell, G. E. (co-authors) : X-rays 1, 228
Binks, W., *see* Kaye, G. W. C.
Birge, T. : The general physical constants 8, 90
Blackman, M. : The theory of the specific heat of solids 8, 11
Bleaney, B. : The radio-frequency spectroscopy of gases 11, 178
de Boer, J. : Molecular distribution and equation of state of gases . . . 12, 305
Born, M. : The teaching of theoretical physics in universities 8, 1
Born, M. : Theoretical investigations of the relation between crystal dynamics and x-ray scattering 9, 294
Bosley, W., and Craggs, J. D. : Nuclear experiments with high voltage x-rays . 12, 82
Boutry, G. A. : Photoelectricity 3, 272
Brown, R. C. (co-author) : General physics 3, 1; 5, 1
Brown, R. C. (co-author) : Surface tension 4, 40; 7, 180
Bruckshaw, J. McG. : Physics and the search for oil 9, 198
Brunt, D. : Meteorology 5, 100
Buckley, H. : Photometry 8, 318
Buckley, H. F. (co-author) : Optics 1, 131
Buechner, W. W., *see* Van de Graaff, R. J.

Cabannes, J. : Spectroscopy applied to molecular physics 10, 378
Cabrera, N., and Mott, N. F. : Theory of the oxidation of metals . . . 12, 163
Callendar, G. S., *see* Sutherland, G. B. B. M.
Cassie, A. B. D. : Physics and textiles 10, 141
Champion, F. C. : The single scattering of elementary particles by matter . . 5, 348
Chapman, S. : Atmospheric physics and chemistry 9, 1
Chapman, S. : The photo-chemistry of atmospheric oxygen 9, 92

Chapman, S., and Price, W. C. : The upper atmosphere 3, 42
Childs, W. H. J. (co-author) : Spectroscopy 3, 369
Christian, R. S. : Analysis of high-energy neutron–proton and proton–proton scattering data 15, 68
Collie, C. H. : Atomic physics 4, 173
Cooke, A. H. : Paramagnetic relaxation effects 13, 276
Cowling, T. G. : The absorption of water vapour in the far infra-red . . . 9, 29
Cox, H. L. (co-author) : General physics 5, 1
Cox, H. L., and Sopwith, D. G. (co-authors) : General physics 3, 1
Craggs, J. D. : Recent research on counter tubes 9, 137
Craggs, J. D., *see* Bosley, W.
Crawford, F. H. (co-author) : Spectroscopy 6, 137
Crawford, M. F. (co-author) : Spectroscopy . . . . . 3, 369 ; 5, 210
Crommelin, A. C. D. : Recent progress in astronomy 5, 82

Deutsch, M. : Angular correlation in nuclear reactions 14, 196
Dingle, H. (co-author) : Spectroscopy . . . . . 1, 102 ; 2, 329

Eirich, F. : Viscosity and the nature of substances of high molecular weight in solution 7, 329
von Engel, A., *see* Lunt, R.

Fage, A. : Fluid motion . . . . . 3, 26 ; 6, 270
Fairweather, A., Roberts, F. F., and Welch, A. J. E. : Magnetic properties of ferrites 15, 142
Feather, N. : Atomic physics . . . . . 2, 74 ; 3, 66
Feather, N. : The gamma radiations emitted in nuclear processes . . . 7, 66
Feather, N. : Radioactive branching 11, 19
Feenberg, E., *see* Wigner, E. P.
Ferguson, A. : Classical physics : surface tension 1, 30
Ferguson, A. : The development of the teaching of experimental physics in British universities 7, 355
Ferguson, A. : General physics (classical). 2, 1
Finch, G. I. (editor) : Electron tubes 2, 258
Flint, H. T. : The beginnings of the new quantum theory 4, 134
Flint, H. T. : Some difficulties and developments in quantum mechanics . . 5, 402
Foster, E. W. : Nuclear effects in atomic spectra 14, 288
Foster, E. W. (co-author) : Spectroscopy 4, 295
Foster, J. Stuart (co-author) : Spectroscopy 5, 210
Francis, V. J., and Jenkins, H. G. : Electrical discharges in gases and their applications 7, 230
Fremlin, J. H., and Gooden, J. S. : Cyclic accelerators 13, 295
Friedenstein, H., Martin, S. L., and Munday, G. L. : The mechanism of the thermionic emission from oxide-coated cathodes 11, 298
Fröhlich, H. : Dielectric breakdown in solids. 6, 411
Fry, D. W., and Walkinshaw, W. : Linear accelerators 12, 102

Garlick, G. F. J. : Phosphors and phosphorescence 12, 34
Garton, W. R. S. (co-author) : Spectroscopy 4, 295
Gassiot Committee Report 9, 1
Gaydon, A. G. : Applications of spectroscopy to combustion 8, 50
Gerson, N. C. : A critical survey of ionospheric temperatures 14, 316
Goldberg, L. : Recent advances in infra-red solar spectroscopy . . . 13, 24
Gooden, J. S., *see* Fremlin, J. H.
Goodeve, C. F. (co-author) : General physics 5, 1
Goodwin, E. T. (co-author) : X-rays and crystals 6, 345
Griffiths, Ezer : Refrigeration 4, 124

Guild, J. (co-author) : Optics **1,** 131; **2,** 356
Gurney, R. W. : Electrolytes and electrolysis **4,** 251
Gurney, R. W., *see* Mott, N. F.

Hampton, W. H., and Wheat, W. N. (co-authors) : Optics **5,** 182
Harrison, G. R. : The M.I.T. wavelength project **8,** 212
Hartree, D. R. : The calculation of atomic structure **11,** 113
Hartshorn, L. : Absolute electrical measurements **5,** 302
Hartshorn, L. : Electric and magnetic measurements **1,** 321
Hartshorn, L. : Electrical and magnetic measurements **2,** 221
Hartshorn, L. : Experimental electricity and magnetism **4,** 218
Hartshorn, L. : Measurements with ultra-short electric waves on wires . . **6,** 378
Hartshorn, L. (co-author) : Experimental electricity and magnetism . . **3,** 206
Heitler, W. : Cosmic rays **5,** 361
Herlofson, N. : The theory of meteor ionization **11,** 455
Hunter, A. : The absorption of light in interstellar space **7,** 150
Hunter, A. : Instrumental technique in astrophysics **8,** 186
Hunter, A. : Solar physics **9,** 101
Hunter, A. : The spectrum of the Sun in the far ultra-violet **9,** 5
Hunter, A. (co-author) : Spectroscopy . **1,** 102; **2,** 329; **3,** 369; **4,** 295; **5,** 210; **6,** 137

Jackson, C. V. (co-author) : Spectroscopy **1,** 102; **2,** 329
Jackson, L. C. : Superconductivity **6,** 335
Jeffreys, Harold : Seismology **10,** 52
Jenkins, H. G., *see* Francis, V. J.
Jenkins, R. O. : Field emission of electrons **9,** 177
Jevons, W. (editor) : Spectroscopy **3,** 369; **4,** 295; **5,** 210; **6,** 137
Johnson, B. K., *see* Martin, L. C.
Jones, E. Gwynne (co-author) : Spectroscopy **1,** 102; **2,** 329
Jones, G. O. : Viscosity and related properties in glass **12,** 133
Jones, H. : The properties of liquid helium **6,** 280
Jones, T. Iorwerth (co-author) : Experimental electricity and magnetism . . **4,** 218

Kaye, G. W. C. : The measurement of noise **3,** 130
Kaye, G. W. C., Bell, G. E., Binks, W., and Perry, W. E. : The production and measurement of short-wave radiations **6,** 95
Kaye, G. W. C. (editor) : X-rays **1,** 288; **2,** 301; **3,** 338
Kellner, L. : Dispersion in the far infra-red **8,** 200
Kellner, L. : Hydrogen bond **15,** 1
Kellner, L. (co-author) : Spectroscopy **2,** 329; **4,** 295
Kerridge, Phyllis M. Tookey : Instrumental aids for defective hearing . . **5,** 150
Kittel, C. : Ultrasonics research and the properties of matter **11,** 205
Klemperer, O. : Electron microscopes of high magnification **7,** 107
Koets, P. : Phase separation in aqueous colloidal systems **10,** 129
Kompfner, R. : Travelling-wave tubes **15,** 275
Kuhn, H. : New techniques in optical interferometry **14,** 64
Kuiper, G. P. : Planetary and satellite atmospheres **13,** 247

Lamb, W. E., Jr. : Anomalous fine structure of hydrogen and singly ionized helium **14,** 19
Lancaster-Jones, E. : Geophysical prospecting **2,** 97
Lee, E., *see* Sutherland, G. B. B. M.
Lee, H. W. : New lens systems **7,** 130
Lipson, H. (co-author) : X-rays and crystals **6,** 345
Lonsdale, Kathleen : Diamagnetic and paramagnetic anisotropy of crystals . . **4,** 368
Lonsdale Kathleen : Experimental study of x-ray scattering in relation to crystal dynamics **9,** 256

Lord, Mary P., and Wright, W. D. : The investigation of eye movements . . 13, 1
Lovell, A. C. B. : Meteoric ionization and ionospheric abnormalities . . . 11, 415
Lovell, A. C. B. : Meteors, comets and meteoric ionization 11, 389
Ludlam, F. H., *see* Mason, B. J.
Lunt, R., von Engel, A., and Meek, J. M. : Discharge phenomena in gases . . 8, 338

McBain, J. W., and Mills, G. F. : The adsorption theorem of J. Willard Gibbs . 5, 30
MacDonald, D. K. C. : Spontaneous fluctuations 12, 56
MacDonald, D. K. C., and Sarginson, K. : Galvanomagnetic effects in conductors 15, 249
McLennan, J. C. : Electrical phenomena at extremely low temperatures . . 1, 198
McVittie, G. C. : The spiral nebulae and the expansion of the universe . . 1, 24
Mann, W. B. : The cyclotron and some of its applications 6, 125
Marshak, R. E., *see* Bethe, H. A.
Martin, L. C. (editor) : Optics 3, 290
Martin, L. C. : Optics 5, 182
Martin, S. L., *see* Friedenstein, H.
Martindale, J. G. : Some applications of physics to the processing of textiles . . 10, 172
Marton, L. : Electron microscopy 10, 204
Mason, B. J., and Ludlam, F. H. : The microphysics of clouds . . . 14, 147
Massey, H. S. W. : Collisions between atoms and molecules at ordinary temperatures 12, 248
Massey, H. S. W., and Bates, D. R. : The properties of neutral and ionized atomic oxygen and their influence on the upper atmosphere 9, 52
May, A. Nunn : The conservation of energy and momentum in elementary processes 3, 89
May, A. Nunn : The mechanism of the Geiger counter 5, 390
Mayne, K. I. : Mass spectrometry 15, 24
Mayneord, W. V. : Some applications of nuclear physics in medicine . . . 14, 366
Mayneord, W. V. : The use of x-rays and γ-rays in medicine 5, 284
Meek, J. M., and Perry, F. R. : The lightning discharge 10, 314
Meek, J. M., *see* Lunt, R.
Meinel, A. B. : The spectrum of the airglow and the aurora 14, 121
Melville, H. W. : Plastics in industrial physics 5, 64
Mendelssohn, K. : Low temperature physics 12, 270
Mendelssohn, K. : Superconductivity 10, 358
Miller, J. L., and Robinson, J. E. L. (co-authors) : Electron tubes . . . 2, 258
Mills, G. F., *see* McBain, J. W.
Moon, P. B. : Slow neutrons 4, 198
Moore, H. : Physics in glass technology 11, 342
Mott, N. F. : Reactions in solids 6, 186
Mott, N. F., *see* Cabrera, N.
Mott, N. F., and Gurney, R. W. : Recent theories of the liquid state . . . 5, 46
Mulliken, R. S., and Ricke, C. A. : Molecular electronic spectra, dispersion and polarization 8, 231
Munday, G. L., *see* Friedenstein, H.

National Physical Laboratory : Optics 1, 131
National Physical Laboratory (editor Kaye, G. W. C.) : X-rays . 1, 228 ; 2, 301 ; 3, 338
Nielsen, H. H. : The infra-red spectra of polyatomic molecules 7, 41

Oliver, D. A. (co-author) : Experimental electricity and magnetism . . . 3, 206
Orowan, E. : Fracture and strength of solids 12, 185
Owen, D. : The teaching of physics in technical institutions 6, 431

Pearse, R. W. B. : Spectroscopic observations on comets, meteors and meteorites . 11, 410
Pearse, R. W. B. : The spectrum of the night sky 9, 42
Pearse, R. W. B. (co-author) : Spectroscopy . 1, 102 ; 2, 329 ; 3, 369 ; 4, 295 ; 5, 210
Peierls, R. : The Bohr theory of nuclear reactions 7, 87

Peierls, R. : The meson **6,** 78
Peierls, R. (co-author) : Quantum theory **2,** 27
Pelzer, H. : Kinetic theory of elasticity of rubber **6,** 330
Penman, H. L. : Evaporation in nature **11,** 366
Penney, W. G. : The theory of molecular structure **6,** 212
Penney, W. G. (co-author) : Quantum theory **2,** 27
Penney, W. G., and Pike, H. H. M. : Shock waves and the propagation of finite pulses in fluids **13,** 46
Penney, W. G., *see* Schlapp, R.
Perry, F. R., *see* Meek, J. M.
Perry, W. E. (co-author) : X-rays **2,** 301; **3,** 338
Perry, W. E., *see* Kaye, G. W. C.
Pike, H. H. M., *see* Penney, W. G.
Pollard, A. F. C. : The mechanical design of physical instruments . . . **10,** 272
Pontecorvo, B. : The neutrino and the recoil of nuclei in beta disintegrations . **11,** 32
Porter, J. G. : Comets and meteor-streams **11,** 402
Powell, C. F. : Mesons **13,** 350
Powell, R. W. : Heat **3,** 143; **4,** 76; **5,** 164; **6,** 297
Prentice, J. P. M. : Meteors **11,** 389
Price, W. C. : Absorption spectra and absorption coefficients of atmospheric gases **9,** 10
Price, W. C. : Recent advances in ultra-violet absorption spectroscopy . . **14,** 1
Price, W. C., *see* Chapman, S.

Randall, J. T. (co-author) : Optics **3,** 290
Randall, J. T., and Wilkins, M. H. F. : The luminiscence of solids . . . **6,** 174
Rawlins, F. Ian G. : The physics of paintings **9,** 334
Raynor, G. V. : Band structure of metals **15,** 173
Richardson, E. G. : Musical acoustics **7,** 27
Richardson, E. G. : Sound **1,** 63; **2,** 166; **3,** 106; **4,** 53; **5,** 121; **6,** 238; **7,** 1; **9,** 228; **10,** 120
Richardson, E. G. : Supersonics in relation to molecular constitution . . **4,** 67
Rideal, E. K. : The absorption of gases by solids **4,** 27
Rieke, C. A., *see* Mulliken, R. S.
Rivlin, R. S. : Impedance networks **6,** 389
Roberts, F. F., *see* Fairweather, A.
Roberts, J. K. : Some interactions of gases with metals and crystalline solids . **7,** 303
Robertson, J. Monteath : X-ray analysis and application of Fourier series methods to molecular structures **4,** 332
Robertson, J. Monteath : X-rays and crystals **6,** 345
Robinson, H. R. : The charge of the electron **2,** 247; **4,** 212
Robinson, J. E. L., *see* Miller, J. L.
Rochester, G. D., and Rosser, W. V. G. : The nuclear interactions of cosmic rays **14,** 227
Rollin, B. V. : Nuclear paramagnetism **12,** 22
Rooksby, H. P. : X-ray diffraction techniques in the industrial laboratory . **10,** 83
Ross, M., and Story, J. S. : Slow neutron absorption cross sections of the elements **12,** 291
Rosser, W. V. G., *see* Rochester, G. D.
Rubinowicz, A. : Multipole radiation in atomic spectra **12,** 233
Ryle, M. : Radio astronomy **13,** 184

Sarginson, K., *see* MacDonald, D. K. C.
Sayers, J. : Ionization phenomena in the earth's atmosphere **9,** 52
Schlapp, R., and Penney, W. G. (co-authors) : Quantum theory . . . **2,** 27
Schurmann, R. : Friction between solid bodies **8,** 71
Selwyn, E. W. H. (co-author) : Optics **5,** 182
Shearer, G. (co-author) : X-rays **1,** 228; **2,** 301; **3,** 338
Shelby, R. E., *see* Zworykin, V. K.

Shenstone, A. G. (co-author): Spectroscopy 5, 210
Shields, R. B., *see* Thode, H. G.
Simeon, F.: Physical-optical instruments and materials 4, 283
Simeon, F. (co-author): Optics 3, 290
Skinner, H. W. B.: The soft x-ray spectroscopy of the solid state . . . 5, 257
Smith, T. (co-author): Optics 1, 131
Smith, T. (editor and co-author): Optics 2, 356
Sopwith, D. G., *see* Cox, H. L.
Spencer Jones, H.: The measurement of time 4, 1
Starks, H. J. H. (co-author): Electron tubes 2, 258
Stiles, W. S. (co-author): Optics 3, 290
Stockmayer, W. H., *see* Beattie, J. A.
Stoner, E. C.: Ferromagnetism 11, 43
Stoner, E. C.: Ferromagnetism: Magnetization curves 13, 83
Story, J. S., *see* Ross, M.
Sutherland, G. B. B. M., and Callendar, G. S.: The infra-red spectra of atmospheric gases other than water vapour 9, 18
Sutherland, G. B. B. M., and Lee, E.: Developments in the infra-red region of the spectrum 11, 144
Sutherland, G. B. B. M. (co-author): Quantum theory 2, 27
Swaine, W. (co-author): Optics 3, 290
Swann, W. F. G.: Cosmic rays 10, 1

Taylor, A. E.: Range–energy relations 15, 49
Taylor, W. H. (co-author): X-rays and crystals 6, 345
Temple, G. (editor): The quantum theory 2, 27
Temple, G.: Quantum and wave mechanics 1, 1
Thackeray, A. D.: Solar physics 7, 160
Thewlis, J. (co-author): X-rays 2, 301; 3, 338
Thode, H. G., and Shields, R. B.: Mass spectroscopy 12, 1
Treloar, L. R. G.: The structure and elasticity of rubber 9, 113
Trubridge, G. F. P.: Burgers' theory of turbulence 1, 43
Trump, J. G., *see* Van de Graaff, R. J.

Urey, H. C.: Separation of isotopes 6, 48

Van de Graaff, R. J., and Buechner, W. W.: Electrostatic generators for the acceleration of charged particles 11, 1

Walke, H. J.: Induced radioactivity 6, 16
Walkinshaw, W., *see* Fry, D. W.
Ward, W. H., (co-author): Experimental electricity and magnetism . . . 4, 218
Welch, A. J. E., *see* Fairweather, A.
Wheat, W. N., *see* Hampton, W. M.
White, H. E., (co-author): Spectroscopy 6, 137
Wigner, E. P.: Symmetry properties of nuclear levels 8, 274
Wilkins, M. H. F., *see* Randall, J. T.
Wilson, A. H.: Superconductivity and the theory of metals 3, 262
Wolf, E.: The diffraction theory of aberrations 14, 95
Wood, W. A., (co-author): X-rays 1, 228; 2, 301; 3, 338
Wright, W. D.: Photo-electric spectrophotometry and its use in industry . . 7, 36
Wright, W. D. (co-author): Optics 3, 290; 5, 182
Wright, W. D., *see* Lord, Mary P.
Wynn-Williams, C. E.: Electrical methods of counting 3, 239

Zworykin, V. K., and Shelby, R. E.: Recent television developments . . . 8, 135